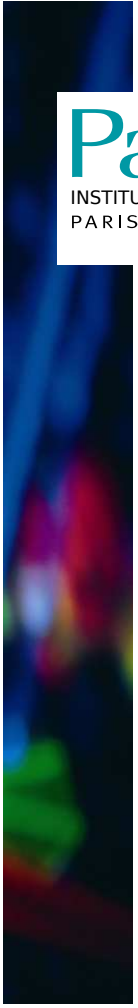
Semantic analysis of 3D point clouds from urban environments : ground, facades, urban objects and accessibility

Andrés Felipe Serna Morales



ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

MINES ParisTech

École doctorale n°432 : Sciences des Métiers de l'Ingénieur

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

l'École nationale supérieure des mines de Paris

Spécialité « Morphologie Mathématique »

présentée et soutenue publiquement par

Andrés Felipe SERNA MORALES

le 16 décembre 2014

Analyse sémantique de nuages de points 3D dans le milieu urbain : sol, façades, objets urbains et accessibilité

Encadrement de thèse : **Beatriz MARCOTEGUI ITURMENDI**
Fernand MEYER

Jury

M. Jean SERRA, Professeur émérite, ESIEE, Université Paris-Est	Président
M. Paul CHECCHIN, Enseignant Chercheur HDR, Université Blaise Pascal, Clermont II	Rapporteur
M. Ferran MARQUES, Professeur, Universitat Politècnica de Catalunya, BARCELONATECH	Rapporteur
M. Jesús ANGULO, Responsable École Doctorale, CMM, MINES ParisTech	Examinateur
M. Raouf BENJEMAA, Directeur Recherche et Développement, Trimble France SAS	Examinateur
M. Philippe JAROSSAY, Chef de la division des plans de voirie, Mairie de Paris	Examinateur
M. Nicolas PAPARODITIS, Directeur Scientifique, IGN France	Examinateur
M. Jorge HERNÁNDEZ, Ingénieur R&T, SAFRAN - CRT Pôle TSI	Invité

MINES ParisTech
PSL★ - Research University
CMM - Centre de Morphologie Mathématique, Mathématiques et Systèmes
35, rue Saint-Honoré, 77305 Fontainebleau, France

T H È S E

Contents

List of Figures

List of Tables

1 Introduction

1.1 Résumé

Dans ce chapitre d'introduction, nous présenterons le contexte de cette thèse sur l'analyse sémantique de données 3D dans le milieu urbain. Nous exposerons ensuite le projet TerraMobilita dans le cadre duquel elle a été développée. Les contributions principales seront annoncées ainsi que les publications scientifiques associées.

1.2 Motivation

Current city maps contain information about roads, sidewalks, facades and urban objects such as lampposts, traffic signs, bollards, trees, among others. Creating and updating these maps is very expensive and time consuming because it is manually carried out by topographers at non-mapped or non-updated locations. Nowadays, several mapping agencies (IGN, 2014b,a), public authorities (Paris, 2014; CASQY, 2014) and private companies (PagesJaunes, 2007; Archivideo, 2014; Cyclomedia, 2014; Earthmine, 2014; 1Spatial, 2014; Geoautomation, 2014; Google, 2014a,b; Trimble, 2014a; Earthmine, 2014) begin to consider justifiable adding 3D information to these urban maps.

Developing 3D maps opens a wide range of applications such as urban planning, cultural heritage documentation, virtual tourism, itinerary planning, marketing, navigation systems and video games. Additionally, in the perspective of a sustainable city and in the framework of new legislation about equality of rights for disabled persons, local authorities are required to execute diagnoses and public works in order to guarantee accessibility to public spaces: sidewalks, bike paths, bus stops, among others (LoiHandicap, 2005; UN, 2007). For these applications, 3D urban scannings are required.

Compared to the first 3D scanning systems 30 years ago, current 3D laser scanners are cheaper, faster, more accurate and provide denser 3D point clouds. For example, depending on the acquisition system resolution, it is possible to get millions of points for a few meters of scanned street. A processing pipeline is required in order to create and update maps from 3D point clouds. It usually consists in transforming points into surfaces or geometric primitives for subsequent analysis. These analyses are usually carried out by manually assisted approaches, leading to time consuming procedures, unsuitable for large scale applications. Object extraction from urban scenes is difficult and tedious, and existing semi-automatic methods may not be sufficiently precise nor robust and exhaustive manual corrections are necessary. In that sense, automatic and accurate methods for 3D urban semantic analysis are required.

This Ph.D. thesis is developed in the framework of TerraMobilita project, which aims at developing new automated methods for 3D urban cartography. Further details on this project are given in the following section.

1.3 TerraMobilita project

> *"3D mapping of roads and urban public space, accessibility and soft mobility"*

TerraMobilita project[1] aims at developing new automated processes to create and update 3D urban maps, with centimeter accuracy, using 3D laser scanning and imagery. Such 3D maps will be used to develop new services and applications for urban space, accessibility and soft mobility.

The project is certified by the clusters Cap Digital and Advancity and it has been selected for funding by FUI11 project call in 2011 and it will finish in 2015. The project brings together 8 partners (4 private companies, 3 public institutions, 1 association and 1 administrative manager), as shown in Figure 1.1 and listed below:

- 1Spatial (http://1spatial.com/fr/), project leader,
- TTS THALES (https://www.thalesgroup.com/en/worldwide/defence/training-simulation),

[1] For further information on our contributions to TerraMobilita project, please visit: http://cmm.ensmp.fr/TerraMobilita/

Figure 1.1: TerraMobilita project brings together 8 partners: 4 private companies (1Spatial, TTS THALES, Trimble Laser Scanning), 3 public institutions (Cityway, IGN, ARMINES/MINES ParisTech), 1 association (CEREMH), 1 administrative manager (Tecdev) and several local authorities from Paris, Saint-Quentin-en-Yvelines and Lille.

- Trimble laser scanning (http://www.trimble.com/3d-laser-scanning/),
- Cityway (http://www.cityway.fr/),
- IGN (http://www.ign.fr/),
- ARMINES/MINES ParisTech - PSL* Research University:
  - CAOR - Center for robotics (http://caor-mines-paristech.fr/)
  - CMM - Center for mathematical morphology (http://cmm.ensmp.fr/)
  - CAS - Center for systems and control (http://cas.ensmp.fr/)
- Sciences Po (http://www.sciencespo.fr/),
- CEREMH (http://www.ceremh.org/),
- TecDev (http://www.tecdev.fr/), administrative management.

The Research and Development (R&D) consortium of the project is coordinated by the three ARMINES/MINES ParisTech laboratories. Thanks to several local authorities associated to the project, prototyping applications and experiments are carried out on three urban areas in France: Paris, Saint-Quentin-en-Yvelines and Lille.

TerraMobilita expected results are twofold: On the one hand, industrial solutions for acquisition, data processing and production of 3D maps of urban roads and public space. On the other hand, solutions for urban management and maintenance as well as applications and services for soft mobility and automatic accessibility diagnoses of the public space. Specific innovations are:

- To build complete urban 3D maps including 3D data and texture information from 3D point clouds and digital images.
- To develop 3D processing algorithms that allow faster, easier, cheaper and more frequent map updates.
- To develop applications and services to manage and maintain public spaces, produce adaptive itineraries for soft mobility, and make automated accessibility diagnoses for different mobilities.

TerraMobilita project falls within the scope of INSPIRE Directive, which establishes a European spatial data infrastructure to ensure interoperability, dissemination, availability, use and reuse of geographic information in Europe (INSPIRE, 2007). Several commercial products and services, mainly Business to Business, will result from the project:

- Automated processing services of laser scanning data and digital imagery.
- 3D modeling and mapping tools for urban environments.
- 3D information services for management and maintenance of the public space, soft mobility, itinerary planning and accessibility diagnoses for disabled people.

Some use cases of TerraMobilita project, directly related to this thesis, are presented below.

1.3.1 Use case AM1: Urban accessibility diagnosis for people with disabilities

TerraMobilita project responds to challenges of the sustainable city taking into account accessibility issues for soft mobility and persons with disabilities under the terms of French law 2005-102 (LoiHandicap, 2005) and United Nations convention (UN, 2007). Our aim is developing 3D maps identifying all potential barriers for a person with disabilities, declaring obstacles into a database and performing accessibility diagnoses of roads and public spaces. Such application will use 3D scannings in order to provide updated urban information, to automate accessibility diagnosis and to offer adaptive urban itineraries for different types of soft mobility. These services can be offered by a private company in the project or directly managed by local authorities

1.3.2 Use case EP1: Automatic parking statistics

Dealing with cars has a particular interest in the framework of TerraMobilita project. The aim of this use case is the automation of parking statistics using 3D multiple daily scans of parking lots. Classically, parking statistics are manually carried out as follows: i) available parking lots are identified on a given urban zone. Different types of parking lots can be found (delivery, disabled parking place, paid, non-dedicated spaces, among others); ii) multiple daily surveys are manually done by operators taking note of license plates of parked cars; iii) parked cars are classified according to type of parking: resident people (parked during the night), short parking (parked less than 1 hour), working people (parked all the day). iv) finally, statistics are computed in order to analyze the user types and the occupancy rate of each parking lot.

In TerraMobilita project, we propose the following pipeline in order to evaluate the potential of an automatic method using 3D laser scanning data: i) segment parked cars on a surveyed zone; ii) compare cars parked in the same zone at different hours. This comparison is required to determine the occupancy duration of each parking slot. iii) finally, compute parking statistics and present results into a Geographical Information System (GIS).

1.3.3 Use case EP2: Degradation of urban furniture

Degradation of urban furniture is of great interest for local authorities. When a damage is reported in some urban object such as a traffic light, a lamppost or a bollard, it is important to detect and to quantify the degradation in order to plan its reparation.

In the framework of TerraMobilita project, we propose the following work-flow to evaluate the potential of an automatic method using 3D laser scanning data: i) segment pole-like objects on a surveyed zone; ii) compare segmented objects with those reported in the 2D urban map. This comparison is required to determine the object presence or absence. iii) compute geometric features such as length, orientation and verticality of each pole-like object in order to evaluate its degradation. iv) finally, integrate results into a GIS for large-scale analysis.

1.4 Contributions of this thesis

As aforementioned, semantic analyses from 3D data are required in order to create and update urban maps. Such analyses are usually carried out by manually assisted approaches, leading to high time consuming rates, unsuitable for large scale applications. In that sense, this thesis introduces automatic methods for urban semantic analysis. Specifically, we focus on a complete 3D urban analysis method including 6 main steps:

Filtering/preprocessing: since geo-referenced laser scanning data are affected by object reflectance, object speed at the acquisition moment, GPS conditions, among others, it is necessary to apply a filtering/preprocessing step in order to reduce outliers, noise and redundant data.

Ground segmentation and accessibility analysis: defining a Digital Terrain Model (DTM) with extra features, such as access ramps geometry, is useful to establish the suitability of a path for a specific mobility type. For example, high curbs should be avoided for a skater or a person in a wheelchair.

Facade segmentation: useful to characterize the front of a building and to define public space boundaries.

Object detection: an object is considered as correctly detected if it is included in the list of object hypotheses, *i.e.* it has not been suppressed by any filtering/preprocessing method and it has not been included as part of the DTM.

Object segmentation: an object is considered as correctly segmented if it has been perfectly isolated as a single object, *i.e.* there is no under-segmentation, and each individual object is entirely inside of only one connected component, *i.e.* there is no over-segmentation.

Object classification: a semantic category is assigned to each segmented object. Each category represents an urban semantic entity. Depending on the application, several classes can be considered, *e.g.* pedestrians, lampposts, traffic signs, benches, cars, garbage containers, bikes, among others. This separation is useful to produce detailed 3D urban maps, to define the best itinerary for a specific mobility type, to produce parking areas maps and to compute parking statistics.

This Ph.D. thesis, entitled "Semantic analysis of 3D point clouds from urban environments: ground, facades, urban objects and accessibility", has been developed at MINES ParisTech in the Center for Mathematical Morphology (CMM) under the supervision of Dr. Beatriz Marcotegui Iturmendi. We aim at developing automatic methods to process 3D point clouds from urban laser scanning. Our methods are based on elevation images, mathematical morphology and supervised learning. Although the processing of 3D urban data has been underway for many years, automatic semantic analysis is still an active research problem. The development of accurate and fast algorithms in this domain is one of the main contributions of the present thesis.

Under previous TerraNumerica project (CapDigital, 2009), several techniques to filter, segment and classify urban objects from 3D point clouds were developed. In the work by Hernández (2009), accurate results were reported using 3D data from mobile laser scanning (MLS) and terrestrial laser scanning (TLS) systems. In particular, that work has been the starting point of this thesis. Current TerraMobilita project (CapDigital, 2014) brings new challenges related to very dense data, high resolution, new application domains (mobility and accessibility) and large-scale processing issues.

In the framework of the present thesis, several methods in the state of the art have been reviewed and their drawbacks have been pointed out. Additionally, more robust and accurate methods have been developed in order to analyze ground, facades, urban objects and accessibility. Our methods have been validated on several public databases in order to get comparative results with the state of the art. Moreover, our methods have been integrated into a large-scale production chain. In that sense, our results can be exported as 3D point clouds for visualization and modeling purposes and as shapefiles for integration in any GIS.

Each chapter of the present manuscript has been written to be self-contained. This document is organized as follows.

Chapter 2 presents an overview on the different laser scanning technologies used in urban environments, in particular MLS and Stop & Go (S&G) mapping systems. Additionally, we present several public 3D databases in the state of the art as well as the acquisition systems and databases developed in the framework of TerraMobilita project.

Chapter 3 presents an overview on the different 3D data structures used to process and to visualize 3D data. Several data structures have been proposed in the state of the art such as elevation images, triangulation, meshing, octrees and k-D trees. The choice of the best data structure is application dependent and it is possible

to combine some of them to get better results in specific tasks such as visualization, filtering, segmentation and classification. In this thesis, most methods work on elevation images, thus their generation takes an important part of that chapter. Moreover, innovative pre-processing techniques are introduced.

In Chapter 4, we propose an automatic and robust method for urban accessibility diagnoses from 3D point clouds. In the first part, our method segments ground and detects urban objects in order to build a 3D obstacle map useful for itinerary planing. In the second part, automatic methods for segmentation, reconnection and characterization of curbs and urban accessibility analysis are developed.

Chapter 5 proposes several automatic methods to segment facades from 3D point clouds. In our experiments, facades are the highest vertical objects on the urban scene and they appear as elongated structures on the elevation image. Thus, we propose several morphological methods based on geometric and geodesic attributes. These methods are useful to segment facades without including objects connected to them such as motorcycles parked next to facades or pedestrians leaning on walls. Additionally, these methods have been proven to be useful in other industrial applications aiming at segmenting elongated objects, as presented in Chapter 7.

Chapter 6 presents a semantic analysis of 3D urban objects based on mathematical morphology and supervised learning. The focus is automatic detection, segmentation and classification of urban objects from 3D laser scanning data. Our automatic method generates object hypotheses as discontinuities on the ground, thus small and isolated regions are eliminated. Then, connected objects are segmented in order to assign a unique identifier to each individual object. Finally, several geometrical and contextual features are computed for each object and classification is carried out using a support vector machine (SVM) approach.

Chapter 7 introduces several methodological contributions to mathematical morphology. We have developed powerful attribute-based operators useful in a wide range of applications such as: attribute controlled reconstruction, adaptive mathematical morphology, feature extraction, filtering and segmentation. Although, the natural application of these methods in the urban semantic analysis is the segmentation of elongated objects such as facades and curbs, we present other uses such as the segmentation of elongated cells in an industrial application.

Finally, Chapter 8 is devoted to discuss advantages and drawbacks of our proposed methods. Moreover, conclusions and perspectives for future works are presented.

1.5 Associated publications

Several contributions of this thesis have already been published in the following papers (sorted by year):

(Serna et al., 2014a) : A. Serna, B. Marcotegui, E. Decencière, T. Baldeweck, A.-M. Pena, S. Brizion. *"Segmentation of elongated objects using attribute profiles and area stability: application to melanocyte segmentation in engineered skin"*. Pattern Recognition Letters. Special Issue on Advances in Mathematical Morphology. Volume 47, October 2014, Pages 172-182.

(Vallet et al., 2014) : B. Vallet, M. Brédif, A. Serna, B. Marcotegui, N. Paparoditis, 2014. *"TerraMobilita/iQmulus Urban Point Cloud Analysis Benchmark"*. Computers & Graphics. (Submitted on September 4, 2014). Pages 1-14.

(Brédif et al., 2014) : M. Brédif, B. Vallet, A. Serna, B. Marcotegui, N. Paparoditis, 2014. *"TerraMobilita/iQmulus urban point cloud classification benchmark"*. In: IQmulus workshop on Processing Large Geospatial Data. iQmulus/TerraMobilita contest. July 8, 2014, Cardiff (UK). Pages 1-6.

(Serna and Marcotegui, 2014) : A. Serna and B. Marcotegui. *"Detection, segmentation and classification of 3D urban objects using mathematical morphology and supervised learning"*. ISPRS Journal of Photogrammetry and Remote Sensing, Volume 93, July 2014, Pages 243-255.

(Serna et al., 2014b) : A. Serna, B. Marcotegui, F. Goulette and J.-E. Deschaud. *"Paris-rue-Madame database: a 3D mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods"*. In proceedings of ICPRAM2014: 3rd International Conference on Pattern Recognition and Methods. March 6-8, 2014, Angers (France). Pages 1-6.

(Serna and Marcotegui, 2013b) : A. Serna and B. Marcotegui. *"Urban accessibility diagnosis from mobile laser scanning data"*. ISPRS Journal of Photogrammetry and Remote Sensing, Volume 84, October 2013, Pages 23-32. **Publication awarded with the U. V. Helava Award for the 2013 best paper in the ISPRS Journal (volumes 75-86)** http://www.isprs.org/society/awards/helava/2013.aspx.

(Serna and Marcotegui, 2013a) : A. Serna, B. Marcotegui. *"Attribute controlled reconstruction and adaptive mathematical morphology"*. In proceedings of ISMM2013: 11th International Symposium on Mathematical Morphology. pp. 205-216. May 27-29, 2013, Uppsala, Sweden.

(Serna et al., 2012) : A. Serna, J. Hernandez, B. Marcotegui. *"Adaptive Parameter Tuning for Morphological Segmentation of Building Facade Images"*. In proceedings of EUSIPCO2012: 20th European Signal Processing Conference, Bucharest, Rumania August 26-31, 2012. pp. 2268-2272, EURASIP 2012.

(Serna and Marcotegui, 2012) : A. Serna and B. Marcotegui, *"Classification 3D d'objets urbains à partir des données terrestres à balayage laser"*. In 35ème journée ISS France. MINES ParisTech. February 2012, Paris, France (In French).

2 Laser scanning technology and 3D data in urban environments

2.1 Résumé

Dans ce chapitre, nous présenterons les technologies existantes dans le domaine de l'acquisition de données 3D. Nous exposerons les technologies utilisées dans le cadre de TerraMobilita. Ensuite, nous présenterons les différentes bases de données 3D disponibles dans l'état de l'art, ainsi que, celles créées par les partenaires du projet et utilisées dans cette thèse.

2.2 Introduction

In the last 30 years, laser scanning technology has been flourishing as surveying technique for the acquisition of geospatial information in outdoor environments. A wide variety of solutions is commercially available, accompanied by many dedicated data acquisition, processing and visualization tools. Due to recent improvements in quality and productivity, this technology is successfully applied to 3D city modeling, digital terrain model generation, forest monitoring, documentation of cultural heritage, among others (Vosselman and Maas, 2010).

The aim of this chapter is to give an overview on the different laser scanning technologies used in urban environments, in particular the mobile laser scanning (MLS) and Stop & Go (S&G) mapping systems. Besides, we present several public 3D databases in the state of the art as well as the acquisition systems and databases developed in the framework of TerraMobilita project.

Several contributions of this chapter have already been published. We have collaborated in the creation, annotation and publication of several 3D urban databases (Serna et al., 2014b; Brédif et al., 2014) as well as in the definition of evaluation protocols using 2D and 3D manual annotations (Serna and Marcotegui, 2013b; Brédif et al., 2014).

This chapter is organized as follows. Section 2.3 discussed technologies used to acquire 3D urban data. Section 2.4 presents acquisition systems used in the framework of TerraMobilita. Section 2.5 describes the available datasets in the state of the art while Section 2.6 presents those acquired in the framework of TerraMobilita project. Finally, Section 2.7 concludes this chapter.

2.3 Laser scanning technology

In order to obtain exploitable 3D data from urban environments, an acquisition system has to face several issues such as resolution, precision in the localization, information management, processing time and storage capacity. In general, there are three methods to acquire 3D urban data: i) passive methods, such as photogrammetry and stereoscopic vision, give the 3D location of specific points or features extracted from the scene. The point density of these methods depends on the texture of the scene (Cramer, 2010; Bulatov et al., 2012; Grigillo and Kanjir, 2012; Gerke and Xiao, 2013); ii) active methods, such as lasers and structured light, give denser data over all scanned surfaces (Goulette et al., 2006b; Lafarge and Mallet, 2012; Paparoditis et al., 2012); and, iii) hybrid methods, exploiting the complementarity between passive and active methods: laser scanning provides the accurate 3D geometry while photogrammetry provides the realistic texture (Sevcik and Studnicka, 2006; Beger et al., 2011; Gerke and Xiao, 2014).

In this work, only laser acquisitions will be considered. Nowadays, LiDAR technology ("light detection and ranging") has been prospering in the remote sensing community thanks to developments such as: Aerial Laser Scanning (ALS), useful for large scale buildings, roads and forests survey; Terrestrial Laser Scanning (TLS), for more detailed but slower urban surveys in outdoor and indoor environments; Mobile Laser Scanning (MLS), less precise than TLS but much more productive since the sensors are mounted on a vehicle; and more recently, Stop and Go (S&G) systems, easily transportable TLS systems making a trade off between precision and productivity. All these laser scanning technologies differ in terms of data capture mode, project size, scanning

mechanisms, point cloud density, acquisition time, accuracy and resolution. In general, the common aspects are the LiDAR principle and the geo-referencing using Global Positioning System (GPS) and Inertial Measurement Units (IMU).

ALS systems (Figure 2.1(a)) have been operating since the 70s and first works reported precisions around 1 m. However, the first operative applications appeared ten years later, in the 80s, thanks to developments in LiDAR, GPS and IMU technologies (Arp et al., 1982; Krabill et al., 1984; Lindengerber, 1989). The current accuracy of these systems is ±10 cm in the Z-axis and ±50 cm in the XY-plane (Vosselman and Maas, 2010).

TLS systems (Figure 2.1(c)) have been operating since the 80s and have currently reached a level of maturity that enable a widespread deployment. It is possible to use several fixed scanners in different locations and to register their individual scans in order to obtain dense and complete 3D point clouds of urban scenarios. TLS accuracy is the best one, less than 10 mm, compared with other laser scanning systems (Vosselman and Maas, 2010). However, acquisition is time consuming. For example, scanning a district or a whole city may take several weeks or even months.

MLS systems (Figure 2.1(b)) are more recent and are prospering due to new applications in urban environments. Indeed, they unlock the productivity problem. However, precision is lower and processing is more complex because laser and image devices must be geo-referenced to the position and orientation of the vehicle. Then, it is necessary to combine several technologies such as GPS, IMU, images, 3D point clouds and data fusion. These systems have centimeter accuracy under good GPS conditions. Recent developments include the VIAPOLIS prototype (Figure 2.2(c)), created by the National French Mapping Agency (IGN), which is a very light electric vehicle that scans from sidewalks, green spaces and open public spaces.

S&G systems (Figure 2.1(c)) are light-weight mobile systems offering a trade-off between productivity and accuracy. These systems are very useful for fast and accurate acquisitions over the sidewalks and in indoors scenarios. Currently, the accuracy of these systems is lower than 10 mm, *i.e.* the accuracy is comparable to that for TLS systems but the acquisition is up to 10 times faster (Trimble, 2014a).

All these systems contain a scanner unit, a global positioning system and an inertial measurement unit. Besides, some of them simultaneously acquire images with digital cameras. The laser is able to record multiple echoes, their reflectance strength (called also intensity) and additional attributes such as time stamp and echo width (Vosselman and Maas, 2010). Figure 2.1 illustrates the acquisition principle in aerial and mobile laser scanning systems.

2.4 TerraMobilita acquisition systems

In the framework of TerraMobilita[1] project, four acquisition systems, a S&G and three MLS, have been developed, as shown in Figure 2.2. One of them, VIAPOLIS system, is a prototype and it is not fully operational yet. Details on the other three systems are presented below.

2.4.1 Stereopolis II

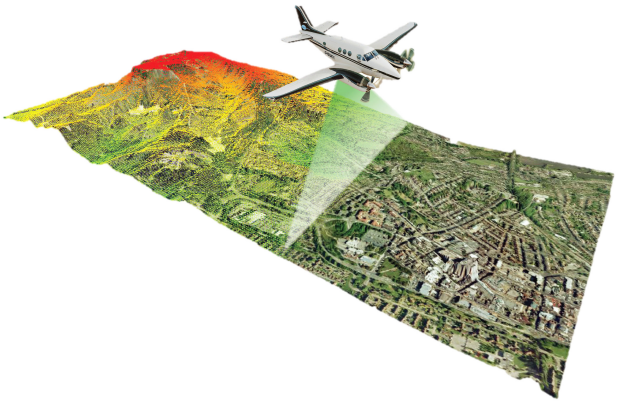
It is a MLS system developed by the National French Mapping Agency (IGN) (Paparoditis et al., 2012), shown in Figure 2.2(a). This system is equipped with a RIEGL VQ-250 laser sensor, generating up to 300,000 points per second with a centimeter resolution. This RIEGL sensor digitizes up to 100 scan lines per second, which gives a spatial sampling of 6 cm along the trajectory direction when the vehicle drives at 20 km/h. Table 2.1 presents the technical specifications of RIEGL VQ-250 laser scanner.

Figure 2.3 presents an example of a 3D dataset acquired by Stereopolis II at *Saint Sulpice square* in Paris, France. This dataset contains 21 million points acquired approximately in 1.5 minutes.

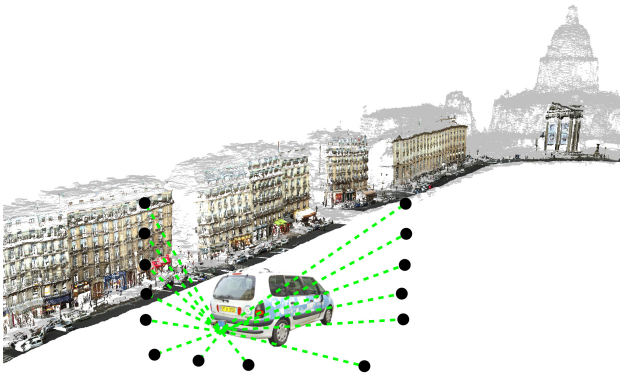
2.4.2 L3D2

It is a MLS system from the robotics laboratory (CAOR) at MINES ParisTech (Goulette et al., 2006b), shown in Figure 2.2(b). This system is equipped with a Velodyne HDL-32E, generating up to 700,000 points per second in a range of 70 meters with a centimeter resolution. In a velodyne sensor, several lasers are mounted on upper

[1] TerraMobilita project: http://cmm.ensmp.fr/TerraMobilita/



(a) Aerial laser scanning (ALS)



(b) Mobile laser scanning (MLS)

(c) Terrestrial laser scanning (TLS) and Stop & Go (S&G). Each color represent a single acquisition.

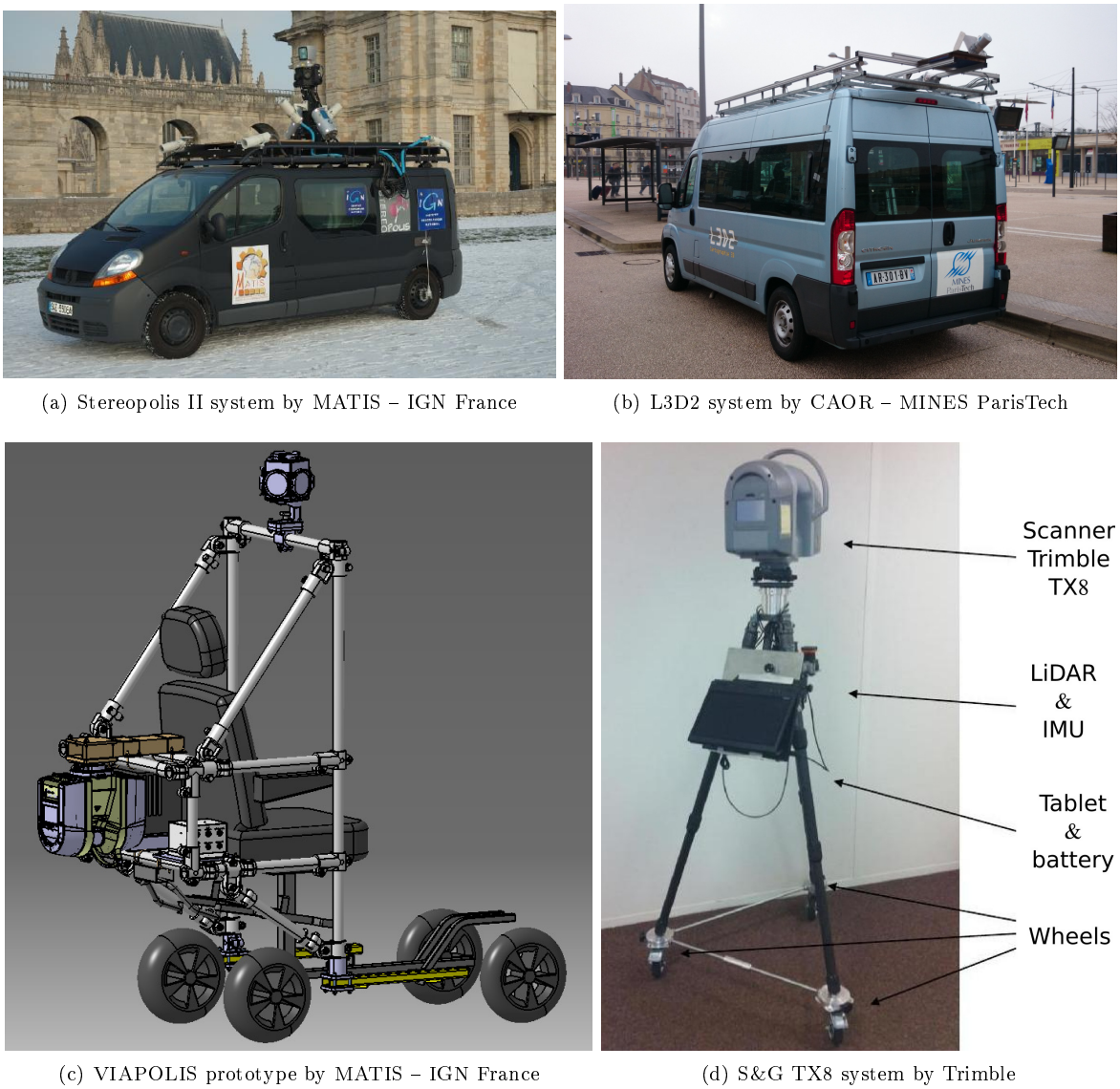
Figure 2.1: Example of acquisition using aerial (ALS), terrestrial (TLS), mobile (TLS) and Stop & Go (S&G) laser systems.

Table 2.1: Technical specifications: RIEGL VQ-250 laser scanner used in Stereopolis II system.

Scanning principle	rotating mirror
Range principle	time of flight measurement
Measurement rate	300 kHz
Minimum Range	1.5 m
Maximum range	75 m
Laser wavelength	near infrared
Vertical field of View	360 degrees
Angular accuracy	350 μrad
Scan Speed	100 scans/sec
Angular resolution	0.001 degrees
Internal Sync Timer	GPS real-time stamping
Accuracy	10 mm
Precision	5 mm
Number of points	300,000 points/sec
Intensity resolution	16 bits

and lower blocks of 32 lasers each and the entire unit spins, giving less precise but much denser point clouds than Riegl sensors (Velodyne, 2012). Table 2.2 presents the technical specifications of Velodyne HDL-32E laser scanner.

Figure 2.4 presents an example of a 3D dataset acquired by L3D2 system in *rue Madame* in Paris, France. This dataset corresponds to a 80 m long street section and contains 10 million points.



(a) Stereopolis II system by MATIS – IGN France

(b) L3D2 system by CAOR – MINES ParisTech

(c) VIAPOLIS prototype by MATIS – IGN France

(d) S&G TX8 system by Trimble

Figure 2.2: MLS and Stop & Go acquisition systems used in the framework of TerraMobilita project.

2.4.3 Stop & Go Trimble TX8

It is a Stop & Go system developed by Trimble Laser Scanning (Trimble, 2014a), shown in Figure 2.2(d). This system uses Trimble TX8 scanner (Trimble, 2014c), allowing 3D spherical acquisitions up to 138 million points for a range of 120 m in less than 3 minutes. Using this system, one can capture detailed datasets at high speed while maintaining high accuracy over the entire range of the scan. Table 2.3 presents the technical specifications of Trimble TX8 laser scanner.

Figure 2.5 presents an example of a 3D dataset acquired by S&G Trimble TX8 system at *Republic square* in Paris, France. This dataset contains 4,000 million points, 40 different scan locations were required and the acquisition time was 4 hours.

2.5 State of the art of 3D databases

Thanks to all the developments on LiDAR technologies, the amount of available 3D geographical data and processing techniques has bloomed in recent years. Many semi-automatic and automatic methods aiming at analyzing 3D urban point clouds can be found in the literature. It is an active research area. However, there

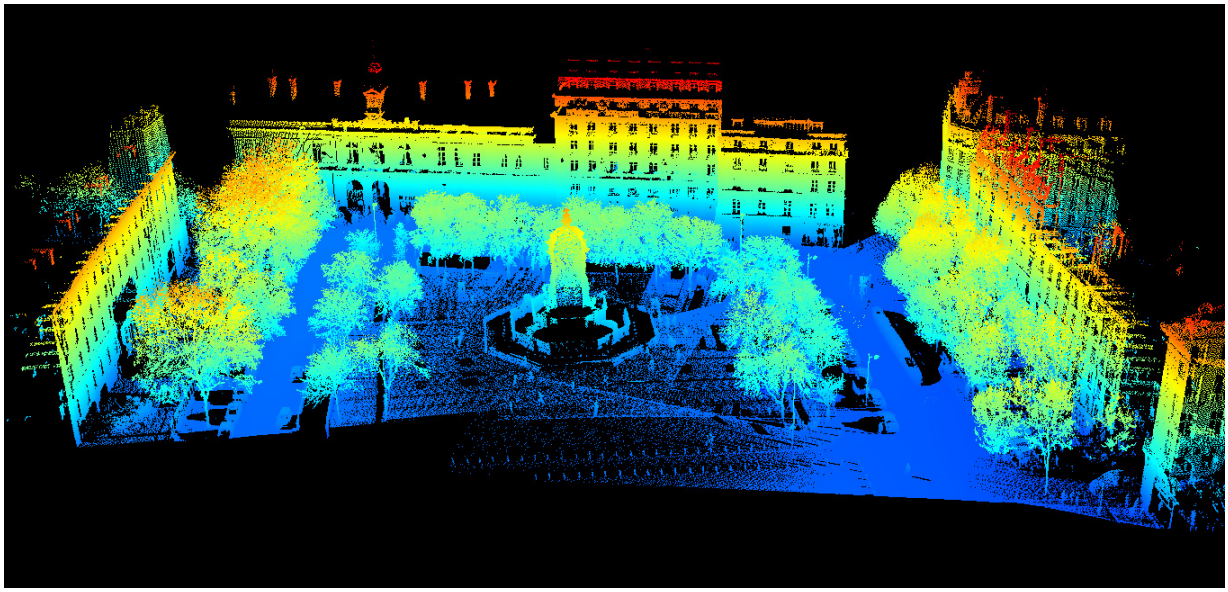


Figure 2.3: 3D point cloud from *Saint Sulpice square* in Paris, France. Acquired by Stereopolis II system, IGN France. The acquisition time was approximately 1.5 minutes.

Table 2.2: Technical specifications: Velodyne HDL-32E laser scanner used in L3D2 system.

Scanning principle	32 lasers on a rotating base
Range principle	time of flight measurement
Measurement rate	10 Hz
Minimum Range	1 m
Maximum range	80 m
Laser wavelength	905 nm
Horizontal field of View	360 degrees
Vertical field of View	$[-30.67, 10.67]$ degrees
Scan Speed	10 scans/sec
Angular resolution	1.33 degrees
Internal Sync Timer	GPS real-time stamping
Accuracy	2 cm
Number of points	700,000 points/sec
Intensity resolution	8 bits

is not a general consensus about the best detection, segmentation and classification methods. This choice is application dependent. One of the main drawbacks is the lack of public databases allowing benchmarking.

In the literature, most available urban data consist in close-range images, aerial images, satellite images but a few laser datasets (ISPRS, 2013; IGN, 2013). Moreover, manual annotations and automatic results are rarely found in available 3D repositories (Nüchter and Lingemann, 2011; CoE LaSR, 2013).

The three following state of the art databases are publicly available and they contain ground truth annotations. They are described here and they will be used to benchmark our methods in the following chapters of this thesis.

2.5.1 Oakland database

Oakland dataset[2] (Munoz et al., 2009) contains 1.6 million points collected around Carnegie Mellon University campus in Oakland, Pittsburgh, USA. Data are provided in ASCII format: (X, Y, Z, label, confidence) one point per line. In this dataset five classes have been annotated: scatter misc, default wires, utility poles, load bearing and facades. Figure 2.6(a) shows a snapshot of this database.

[2]Oakland dataset available at: http://www.cs.cmu.edu/~vmr/datasets/oakland_3d/cvpr09/doc/

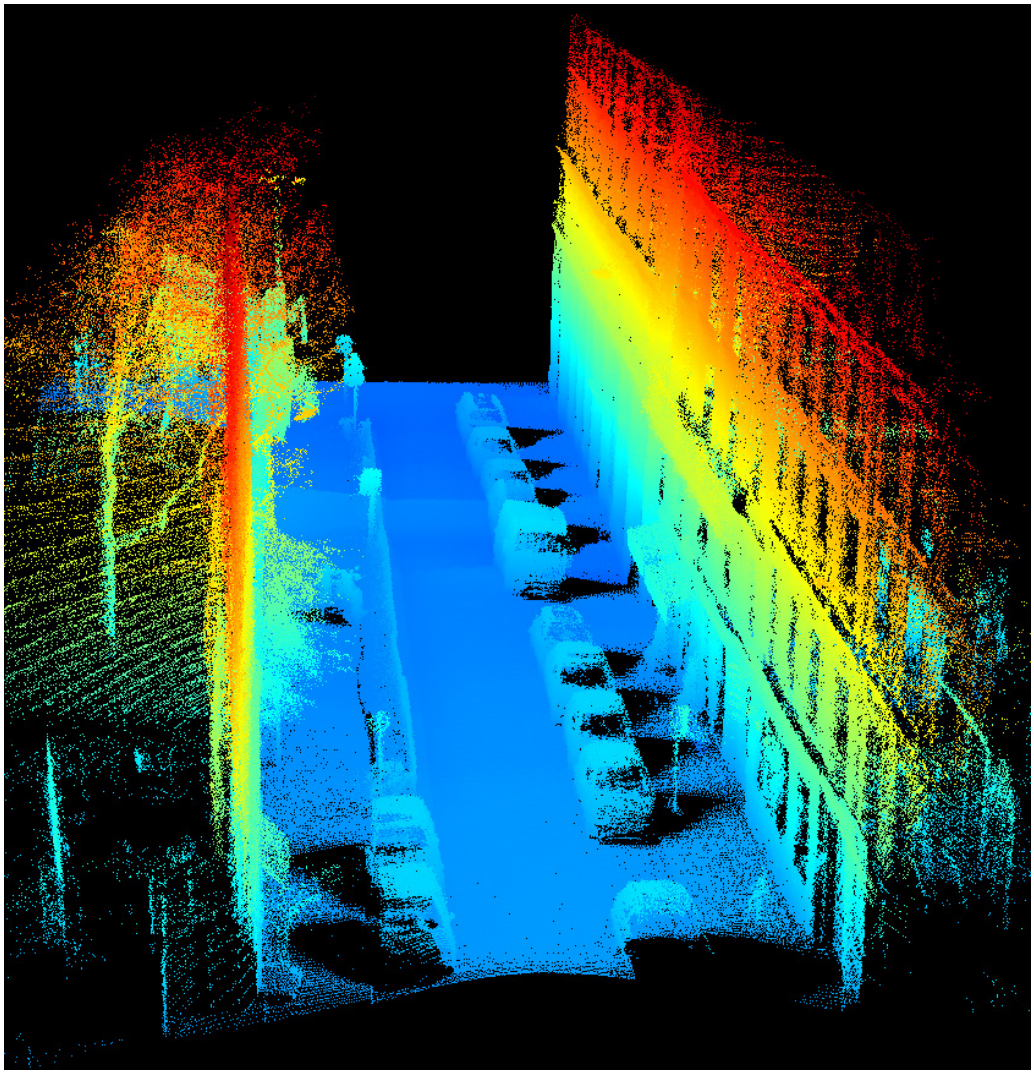


Figure 2.4: 3D point cloud from *rue Madame* in Paris, France. Acquired by L3D2 system by MINES ParisTech. This dataset corresponds to a 80 m long street section and contains 10 million points.

Table 2.3: Technical specifications: Trimble TX8 laser scanner used in the Stop & Go system.

Scanning principle	Vertically rotating mirror on horizontally rotating base
Range principle	Ultra-high speed time of flight
Measurement rate	1 MHz
Minimum range	0.6 m
Maximum range	120 m
Range noise	< 2 mm
Laser wavelength	1.5 μm, invisible
Horizontal field of view	360 degrees
Vertical field of view	317 degrees
Angular accuracy	80 μrad
Scan duration	< 3 minutes
Point spacing at 30 m	22.6 mm
Mirror rotating speed	60 rps
Number of points	1 Mpoints/sec (138 Mpoints in total)
Intensity resolution	8 bits

(a) 3D point cloud from *Republic square* in Paris.

Figure 2.5: 3D point cloud from *Republic square* in Paris, France. Acquired by S&G TX8 system by Trimble Laser Scanning. For this acquisition, 40 different scan locations were required for an acquisition time of 4 hours. This dataset contains 4,000 million points.

2.5.2 Paris-rue-Soufflot database

Paris-rue-Soufflot dataset[3] (Hernández and Marcotegui, 2009a) contains MLS data, acquired by IGN, from a 500 m long street in the 5^{th} Parisian district. Six classes have been annotated: facades, ground, cars, lampposts, pedestrians and others. This database has been created in the framework of TerraNumerica project[4]. It has been used before by Hernández and Marcotegui (2009c) and Serna and Marcotegui (2014) and it will be used later in Chapter 6 in order to benchmark our object classification method. Table 2.4 presents available classes and number of objects by category in Paris-rue-Soufflot dataset. Figure 2.6(b) shows a snapshot of this dataset.

Table 2.4: Available classes and number of objects in Paris-rue-Soufflot database.

Class	Class name	Samples
1	Cars	27
2	Lampposts	12
3	Bollards	39
4	Walls	12
5	Fences	5
6	Pedestrians	101
7	Bikes	14
8	Furniture	30
9	Others	23
10	Traffic lights	4
11	Panels	7
12	Trash cans	5
	Total	**279**

[3] Paris-rue-Soufflot database is available at: http://cmm.ensmp.fr/~serna/downloads.html
[4] TerraNumerica project: http://cmm.ensmp.fr/TerraNumerica/terranumerica.html

2.5.3 Ohio database

Ohio dataset[5] (Golovinskiy et al., 2009) is a combination of ALS and TLS data, acquired by Neptec Technologies Corp (?), in Ottawa city (Ohio, USA). It contains 26 tiles (100 × 100 meters each) with several objects such as buildings, trees, cars and lampposts. However, ground truth annotations only consist in a 2D labeled point in the center of each object. In that sense, segmentation results cannot be evaluated point by point. This database has been used before by several authors (Golovinskiy et al., 2009; Velizhev et al., 2012; Serna and Marcotegui, 2014) and it will be used later in Chapter 6 in order to benchmark our object segmentation and classification methods. Table 2.5 presents available classes and number of objects by category in Ohio dataset. Figure 2.6(c) shows a snapshot of this dataset.

Table 2.5: Available classes and number of objects in Ohio database.

Class	Class name	Samples
1	Ad cylinder	6
2	Bush	29
3	Car	240
4	Dumpster	1
5	Fire hydrant	19
6	Flagpole	2
7	Lamppost	146
8	Light pole	62
9	Mailing box	4
10	Newspaper box	42
11	Parking meter	10
12	Post	377
13	Recycle bin	6
14	Sign	96
15	Telephone booth	4
16	Traffic control box	8
17	Traffic light	42
18	Trash can	19
19	Tree	552
20	Box transformer	2
	Total	**1667**

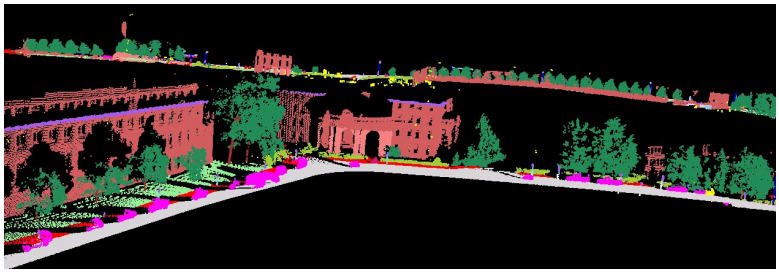
2.5.4 Enschede database

Enschede dataset[6] (Zhou and Vosselman, 2012) is a combination of ALS and MLS data, acquired by FLIMAP (?), from a residential neighborhood approximately 1 km long in Enschede city (The Netherlands). Ground truth annotation consists in 2D geo-referenced lines marking curbstones. A well-defined evaluation method is available using buffers around each 2D line. The drawback of this dataset is that no other objects are annotated. This database has been used before by several authors (Vosselman and Zhou, 2009; Zhou and Vosselman, 2012; Serna and Marcotegui, 2013b) and it will be used later in Chapter 4 in order to benchmark our curb segmentation method. Figure 2.6(d) shows a snapshot of this dataset.

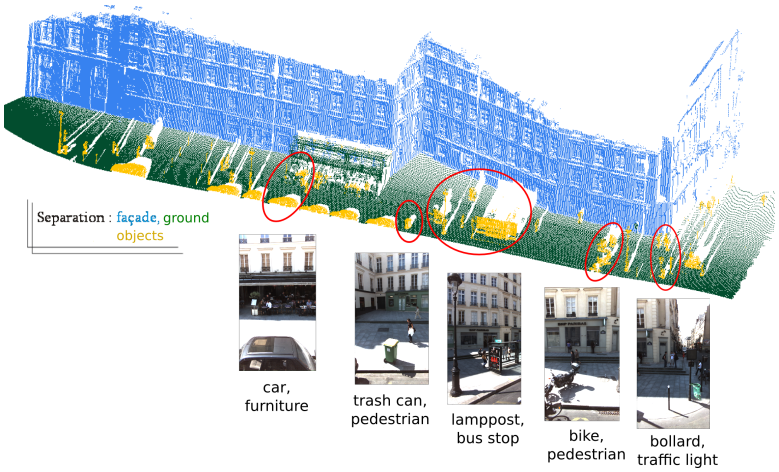
2.6 TerraMobilita 3D databases

In the framework of TerraMobilita project, several databases have been developed in order to set up, test and benchmark semantic analysis methods working on 3D urban data. In this section, we present a description of these databases. For further information on these and other available 3D urban databases, the reader is encouraged to visit: http://cmm.ensmp.fr/~serna/downloads.html

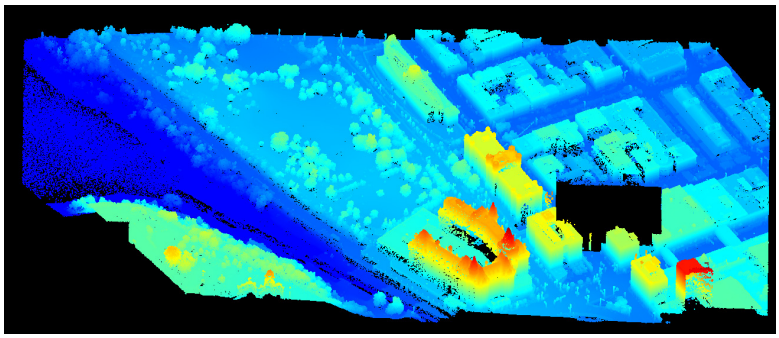
[5] To download Ohio database please contact Dr. Alexander Velizhev: http://graphics.cs.msu.ru/en/node/710
[6] To download Enschede database please contact Prof. Georges Vosselman: http://www.itc.nl/personal/vosselman/



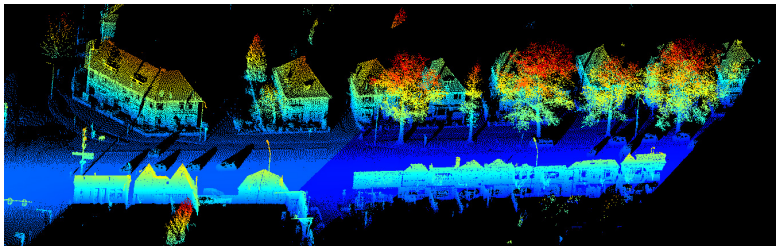
(a) Oakland dataset (MLS)



(b) Paris-rue-Soufflot dataset (MLS)



(c) Ohio dataset (ALS and TLS)



(d) Enschede dataset (ALS and MLS)

Figure 2.6: 3D urban databases in the state of the art. These databases are publicly available and they contain ground truth annotations. They are described here and they will be used to benchmark our methods in the following chapters of this thesis.

2.6.1 Paris-rue-Madame database

Paris-rue-Madame database[7] (Serna et al., 2014b) contains 3D MLS data from *rue Madame*, a street in the 6^{th} Parisian district. Data have been acquired by L3D2 system (Section 2.4.2) in February 2013. Figure 2.7 shows an orthophoto from this test zone, approximately a 160 m long street section between *rue Mézières* and *rue Vaugirard*.

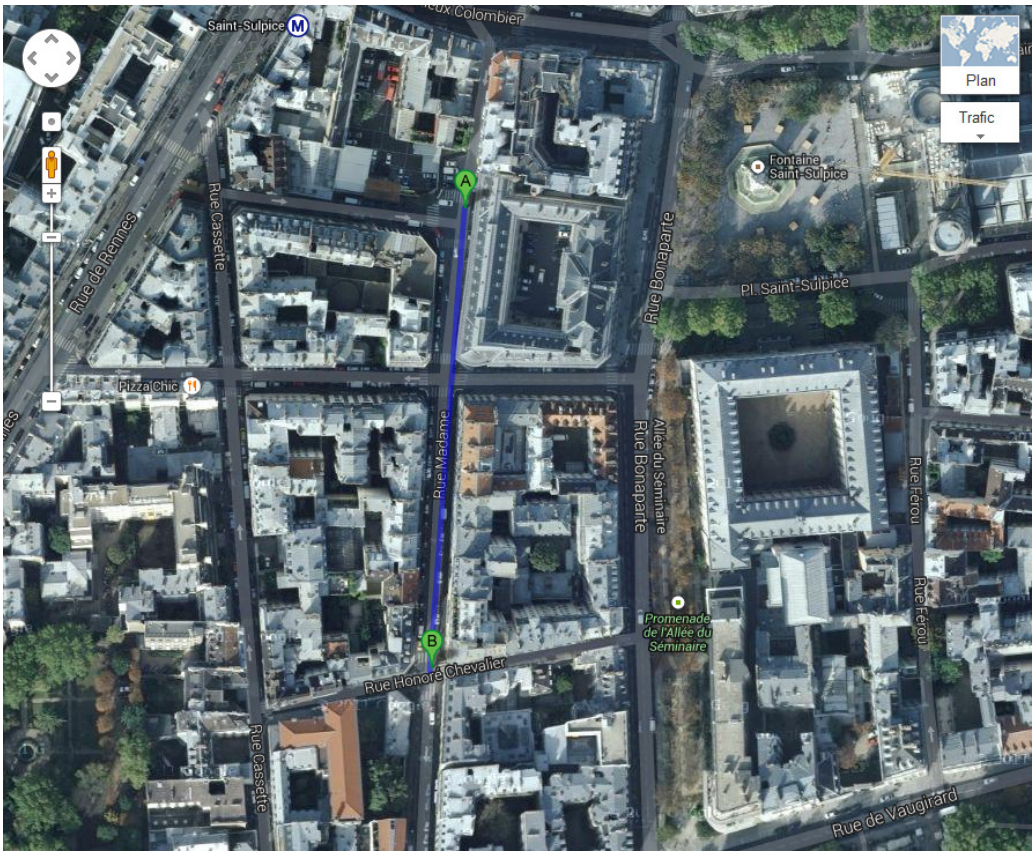


Figure 2.7: Orthophoto from *rue Madame* in Paris, France. Approximately a 160 m long street section. Data have been acquired by L3D2 system (Section 2.4.2) in February 2013. Orthophoto from IGN-Google Maps.

The dataset contains two PLY files with 10 million points each. The available files are "GT_Madame1_2.ply" and "GT_Madame1_3.ply", both of them coded as binary little endian version 1. All coordinates are geo-referenced (E,N,U) in Lambert 93 and altitude IGN1969 (grid RAF09) reference system. An offset has been subtracted from XY coordinates with the aim of increasing data precision: $X_0 = 650976$ m and $Y_0 = 6861466$ m, respectively. Each file contains the following attributes:

(float32) X,Y,Z: Cartesian geo-referenced coordinates in Lambert 93 system.

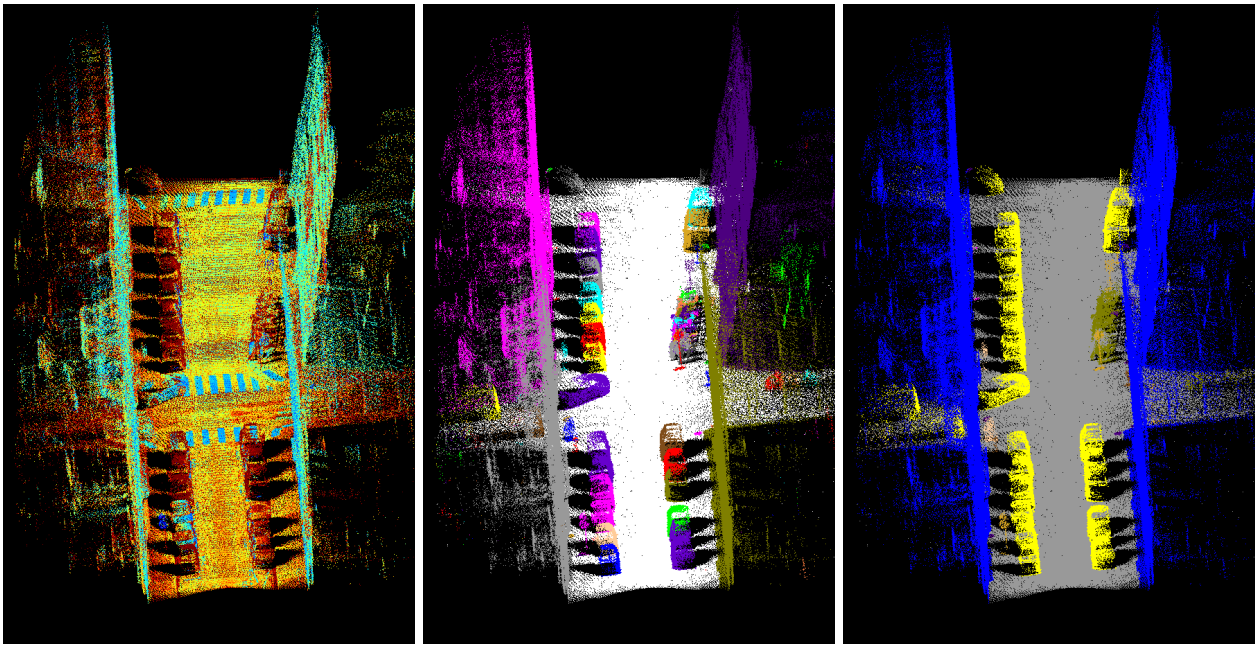
(float32) reflectance: backscattered intensity corrected for distance.

(uint32) id: containing a unique identifier/label for each segmented object.

(uint32) class: containing the classification result for each segmented object. Two points having the same *id* must have the same *class*.

Figure 2.8 presents one of the 3D point clouds of this database colored by the reflectance, the object *id* and the object *class*.

[7] Paris-rue-Madame database is available at: http://cmm.ensmp.fr/~serna/rueMadameDataset.html



(a) Laser reflectance. (b) Object identifier (*id*). (c) Object *class*.

Figure 2.8: Paris-rue-Madame dataset: "GT_Madame1_2.ply" manually annotated file. 3D point cloud colored by its available fields. For object id: each color represents a different object (some colors may look similar when displaying). For object class: facades (blue), ground (gray), cars (yellow), motorcycles (olive), traffic signs (goldenrod), pedestrians (pink).

This database contains 327 objects categorized in 26 classes, as shown in Table 2.6. It is noteworthy that several objects inside buildings have been acquired through windows and open doors. These objects have been annotated as facades.

The annotation is voluntarily very detailed aiming at producing a ground truth useful to a wide range of applications. For example, the *pedestrian* class has been separated in *still pedestrians*, *fast pedestrians* and *pedestrians+something* because all of them have different geometrical features. These subclasses may be gathered or eliminated according to the type of classification we want to evaluate.

It is noteworthy that the entire 3D point cloud has been segmented and classified, *i.e.* each point contains an *id* and a *class*. Thus, the point-wise evaluation of detection, segmentation and classification methods becomes possible.

2.6.2 TerraMobilita/iQmulus database

TerraMobilita/iQmulus database[8] (Brédif et al., 2014) has been developed aiming at benchmarking segmentation and classification methods working on 3D dense urban data. This database has been created in the framework of TerraMobilita project. It consists in 11 annotated test zones containing 30 million points each and 1 annotated training zone containing 12 million points. It has been acquired by Stereopolis II in Paris in January 2013. Annotation has been carried out in a manually assisted way by MATIS laboratory at IGN.

The dataset is presented in PLY format with little endian encoding. All coordinates are geo-referenced (E,N,U) in Lambert 93 and altitude IGN1969 (grid RAF09) reference system. Offsets have been subtracted from XY coordinates with the aim of increasing data precision: $X_0 = 649000$ m and $Y_0 = 6840000$ m, respectively. Each file contains the following attributes:

(float32) X,Y,Z: Cartesian geo-referenced coordinates in Lambert 93 system.

(float32) X,Y,Z origin: Cartesian geo-referenced coordinates of the sensor.

[8] Available at: http://data.ign.fr/benchmarks/UrbanAnalysis/

Table 2.6: Available classes and number of objects in Paris-rue-Madame database.

Class	Class name	Samples file 1_2	Samples file 1_3
0	Background	7	35
1	Facade	4	4
2	Ground	1	1
4	Cars	39	31
7	Light poles	0	1
9	Still pedestrians	3	7
10	Motorcycles	23	9
14	Traffic signs	5	1
15	Trash can	2	1
19	Wall Light	6	1
20	Balcony Plant	3	2
21	Parking meter	1	1
22	Fast pedestrian	2	2
23	Wall Sign	1	3
24	Pedestrian + something	1	0
25	Noise	46	80
26	Pot plant	0	4
	Total	**144**	**183**

(float32) reflectance: backscattered intensity corrected for distance.

(uint8) num_echo: number of the echo (to handle multiple echoes).

(uint32) id: containing a unique identifier/label for each segmented object.

(uint32) class: containing the classification result for each segmented object. Two points having the same *id* must have the same *class*.

In this database, the entire 3D point cloud is segmented and classified, *i.e.* each point contains an *id* and a *class*. Thus, the point-wise evaluation of detection, segmentation and classification methods becomes possible. Figure 2.9 presents one of the 3D point clouds of this database colored by the reflectance, the object *id* and the object *class*.

In this contest, a hierarchy of semantic classes has been defined, as shown in Figure 2.10. The tree is voluntarily very detailed as the aim at producing a ground truth that can be useful to a wide range of methods. The total number of available classes is 101.

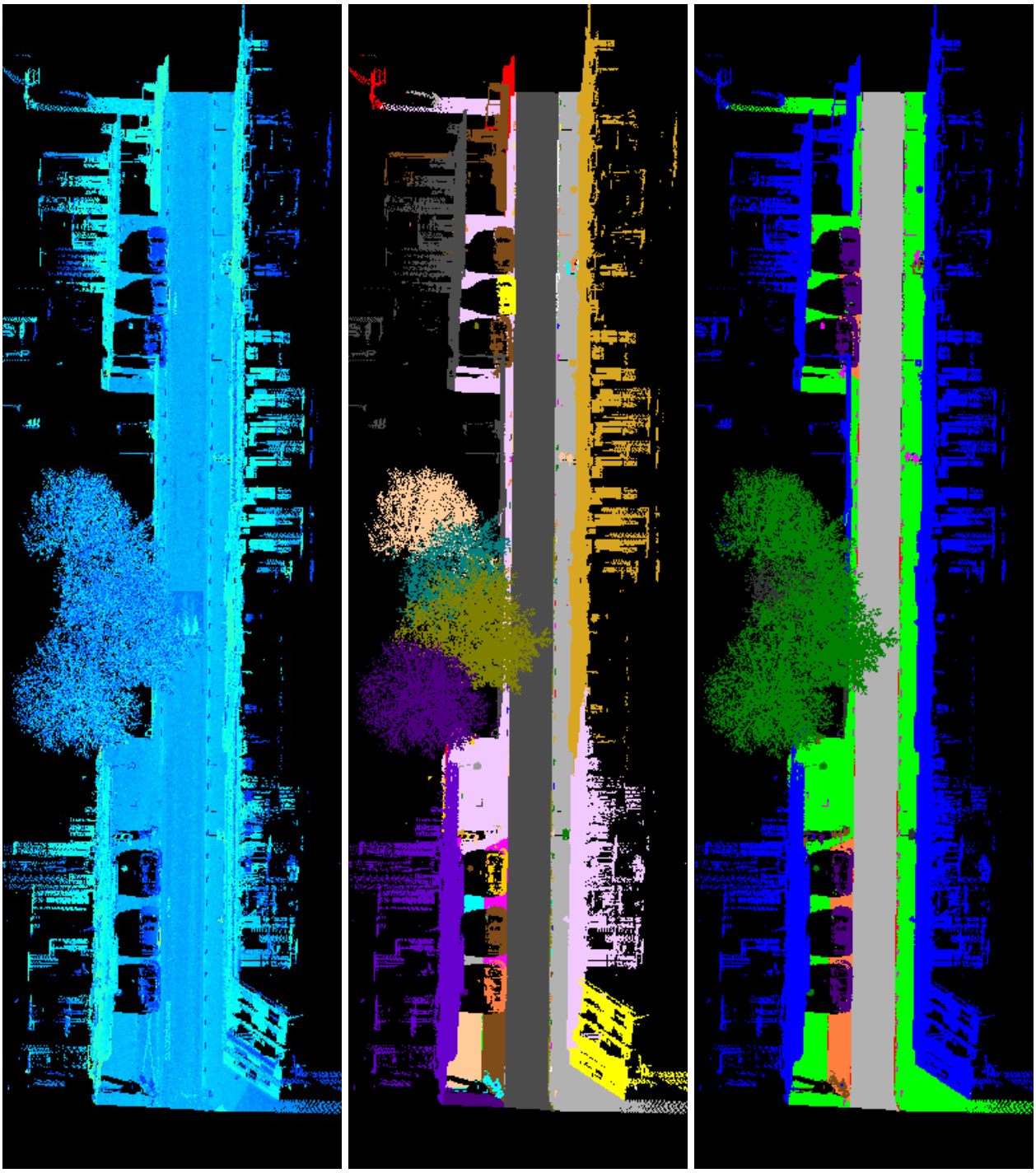
2.6.3 Non-annotated TerraMobilita datasets

Other non annotated datasets have been acquired in the framework of TerraMobilita project. These databases have been acquired for specific aims inside the project (Section 1.3) such as the development of automatic methods for ground coating analysis, parking statistics, urban furniture change detection, documentation of cultural heritage, among others. In the future, some of these datasets will be annotated and made available to the scientific community. For further information the reader is encouraged to visit:
http://cmm.ensmp.fr/~serna/downloads.html

2.6.3.1 MINES ParisTech acquisitions

Figure 2.11(a) presents an experimental zone approximately 2 km long in the 6^{th} Parisian district acquired by L3D2 system (Section 2.4.2) on June 17th, 2014. These data have been specially acquired for the development of an automatic system to compute parking statistics (use case EP1 in Section 1.3). For this, 3D acquisitions of the same parking zone have been carried out at 7 different moments in a day: 11:30 am, 12:30 pm, 1:30 pm, 2:30 pm, 4:00 pm, 6:00 pm and 7:00 pm, respectively.

GPS and IMU data have been post-treated using the RTK basis. The trajectory of each acquisition is provided in an additional point cloud with constant Z = 38 m in order to simplify the comparison of vehicles parked between successive passages.



(a) Laser reflectance. (b) Object identifier (*id*). (c) Object *class*.

Figure 2.9: TerraMobilita/iQmulus database: "Cassette_idclass.ply" manually annotated file. 3D point cloud colored by its available fields. For object id: each color represents a different object (some colors may look similar when displaying). For object class: facades (blue), road (gray), sidewalk (green), other ground (orange), curbs (red), cars (magenta), motorcycles (teal), trees (dark green), undefined (dark gray).

Data is presented in PLY format using ASCII encoding. Each PLY file contains 5 million XYZ+Intensity points. All coordinates have float precision and are geo-referenced (E, N, U) in the Lambert 93 reference

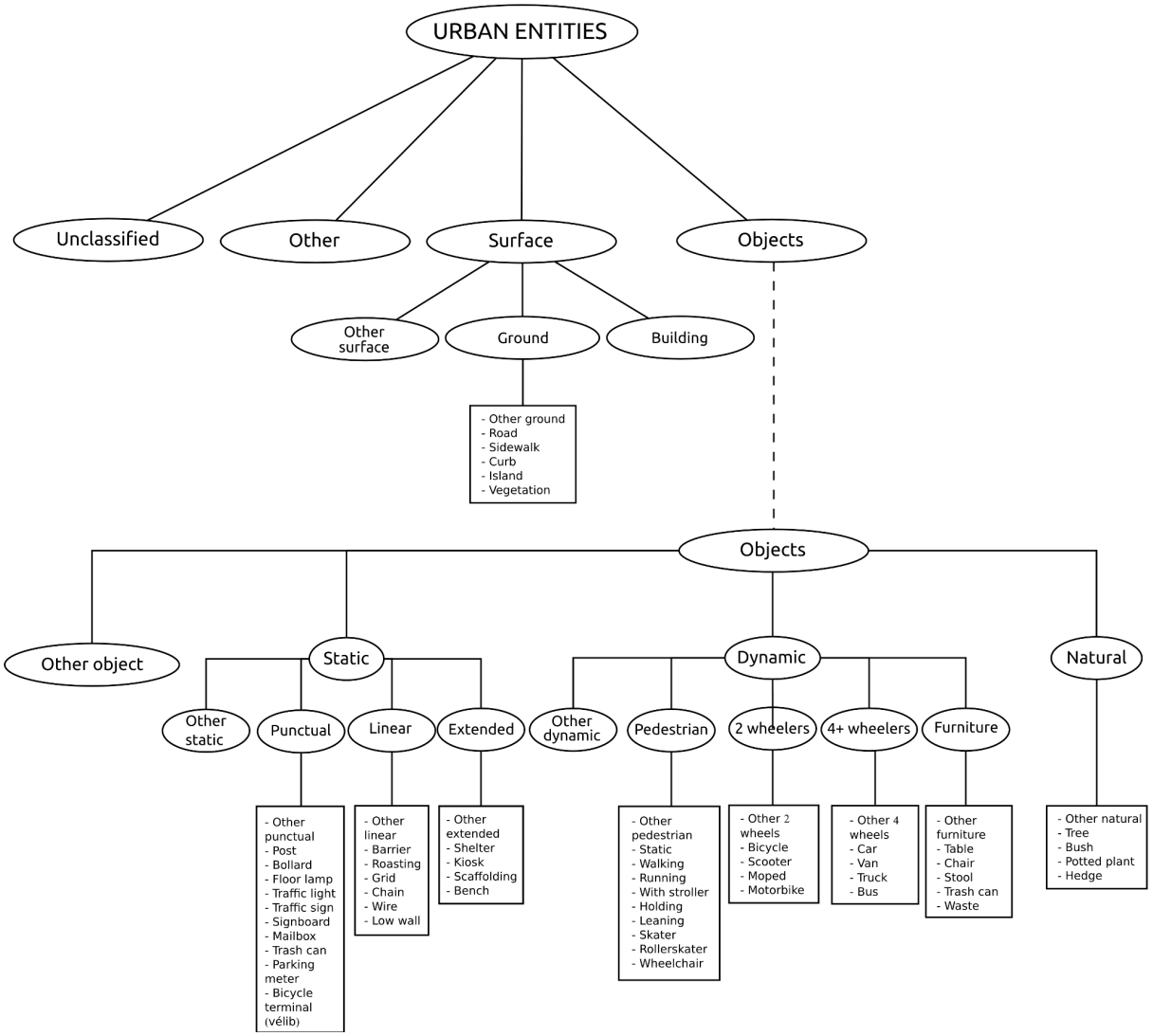


Figure 2.10: Hierarchy of semantic classes defined in TerraMobilita/iQmulus database. The tree is voluntarily very detailed as the aim at producing a ground truth that can be useful to a wide range of methods. The total number of available classes is 101.

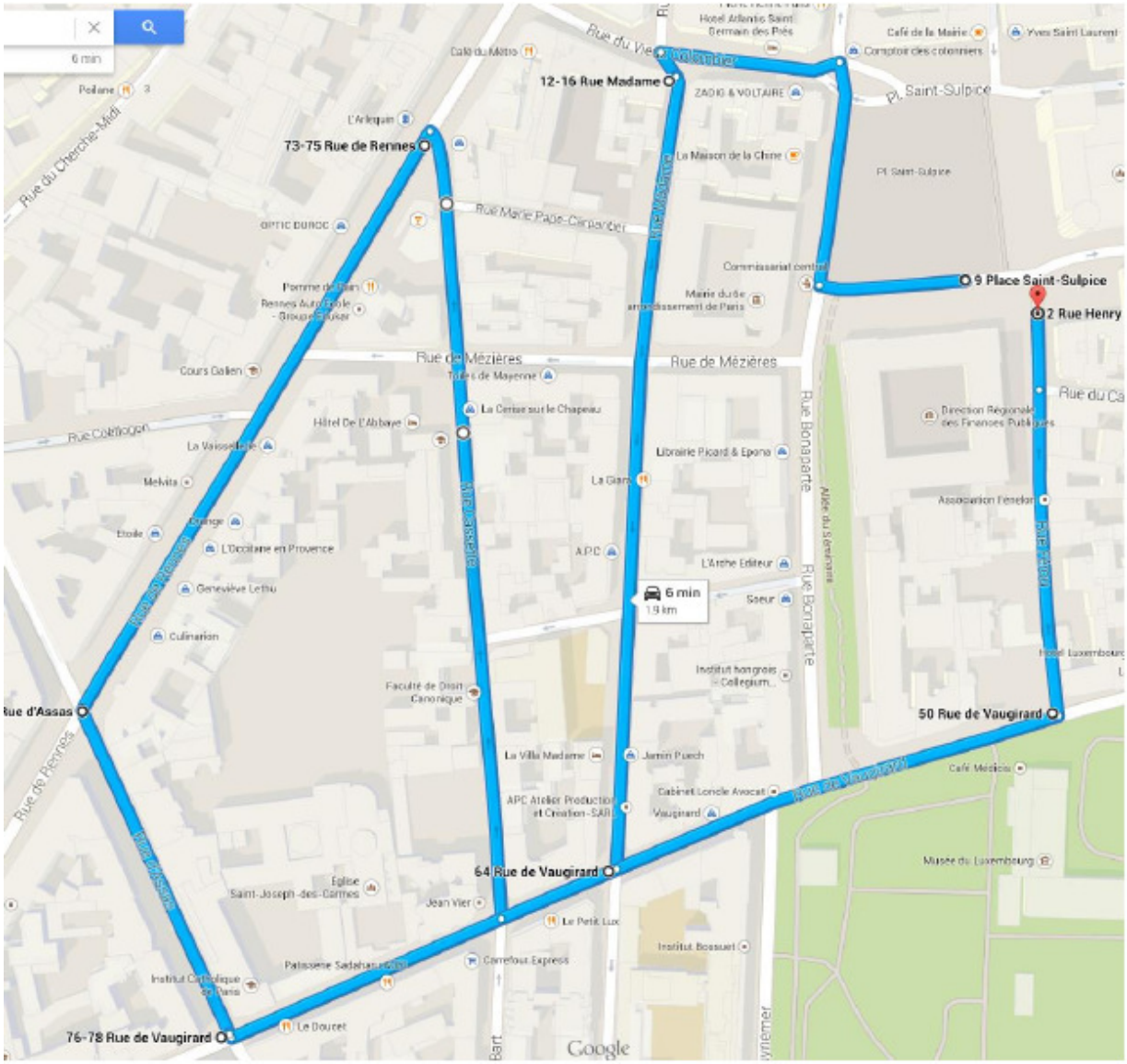
system. The laser intensity is coded as an integer value between 0 and 255. Offsets have been subtracted from XY coordinates in order to avoid loss of information: $X_0 = 650976.0$ m and $Y_0 = 6861466.0$ m, respectively.

Additionally, a video has been recorded in order to make easier the human interpretation of the scene. For this purpose, the camera Garmin Elite Virb has been used. It has been positioned in the middle of the front windshield and it acquires 1920×1080 pixels at 30 frames per second.

Figure 2.11 shows an example of this acquisition, a photo and its corresponding 3D point cloud, in *rue Madame* in Paris, France.

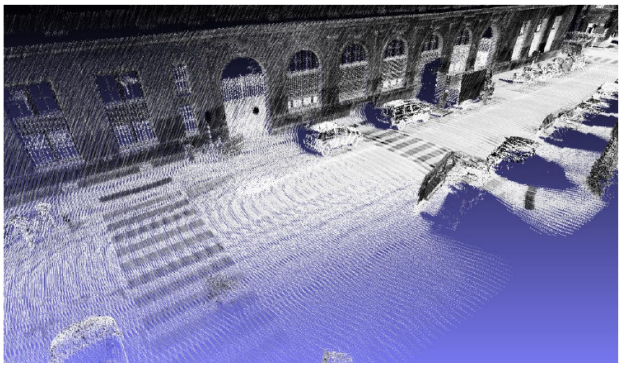
2.6.3.2 IGN acquisitions

Figure 2.12 presents two experimental zones in the 6^{th} Parisian district acquired by Stereopolis II system (Section 2.4.1) in 2012 and 2013, respectively. These data are used with the aim of developing and testing detection, segmentation and classification methods of urban objects. Points properties (origin, reflectance, num_echo, etc.) are the same than in TerraMobilita/iQmulus database (Section 2.6.2).



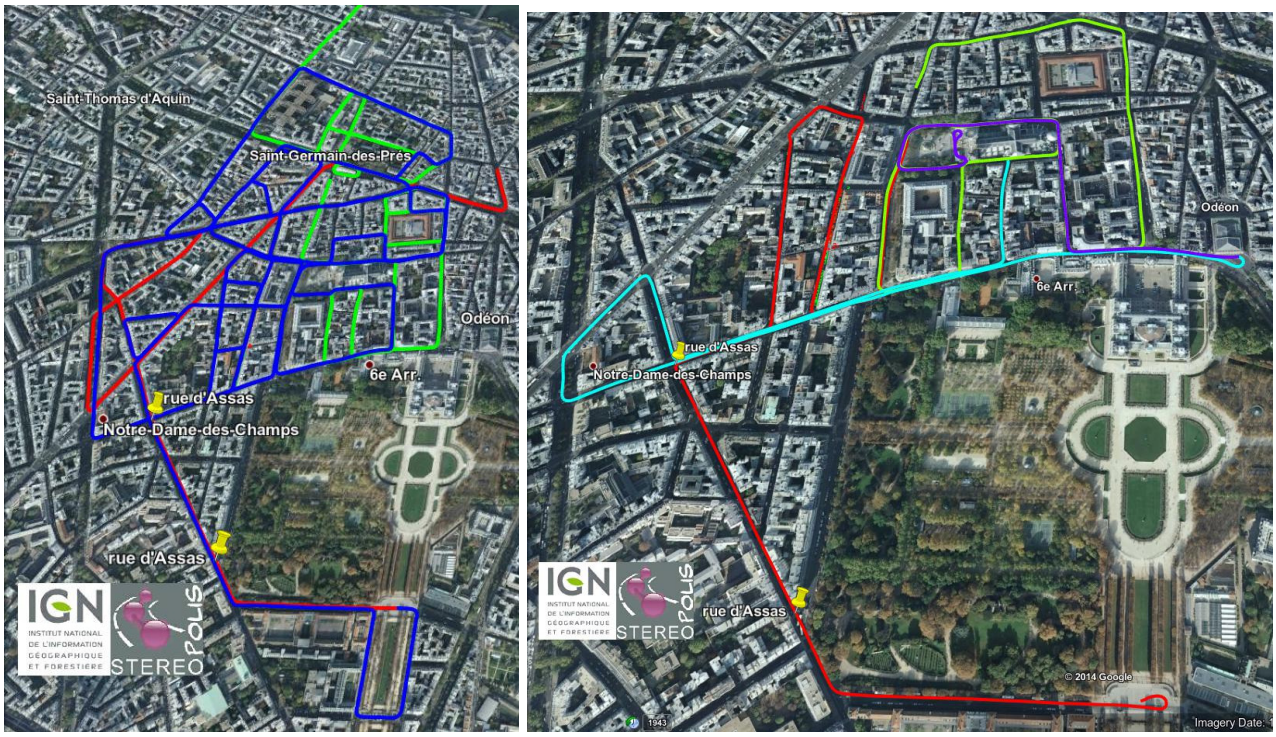
(a) Experimental zone in the 6^{th} Parisian district, France.

(b) Photo in *rue Madame*.



(c) 3D point cloud in *rue Madame*.

Figure 2.11: Example of a non-annotated acquisition by L3D2 system, CAOR-MINES ParisTech. Experimental zone in the 6^{th} Parisian district, France. Map and itinerary taken from Google Maps.



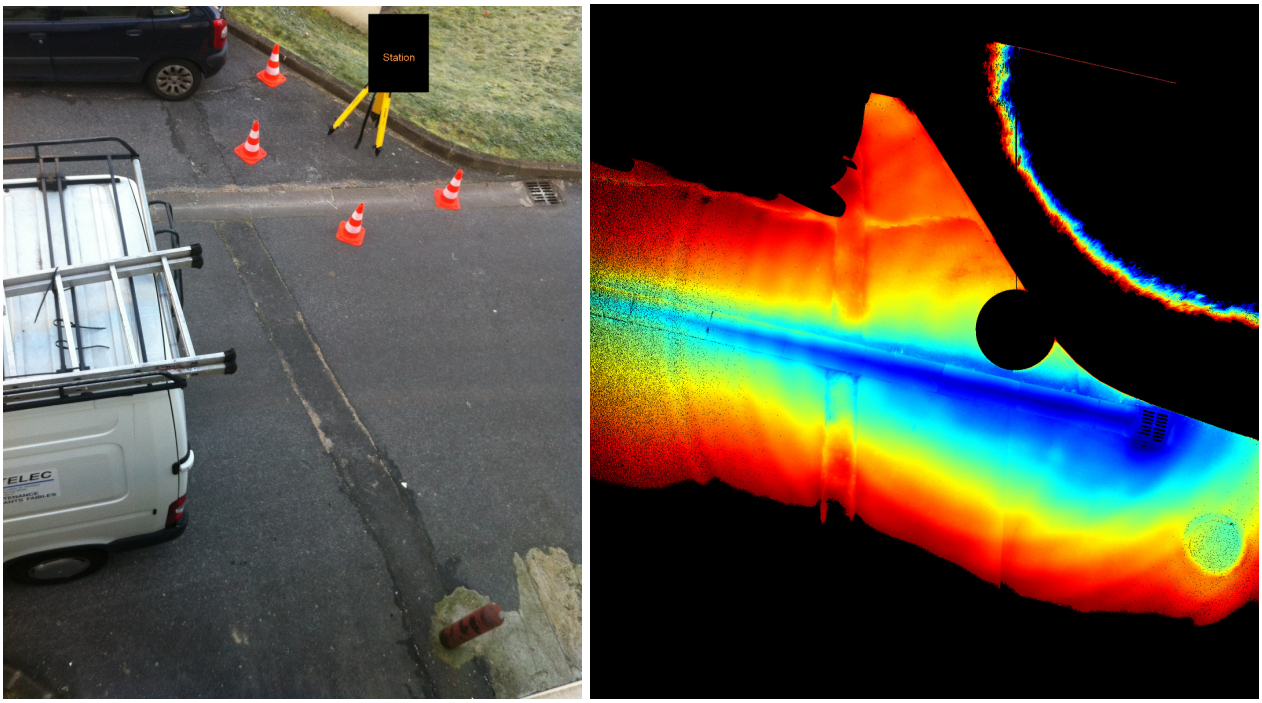
(a) Stereopolis II experimental zone 2012.

(b) Stereopolis II experimental zone 2013.

Figure 2.12: Two experimental zones in the 6^{th} Parisian district acquired by Stereopolis II system (Section 2.4.1) in 2012 and 2013, respectively. These data are used with the aim of developing and testing detection, segmentation and classification methods of urban objects. Points properties (origin, reflectance, num_echo, etc.) are the same than in TerraMobilita/iQmulus database (Section 2.6.2).

2.6.3.3 Trimble acquisitions

Other non-annotated acquisitions have been carried out by Stop & Go Trimble TX8 system (Section 2.4.3). These data have been acquired with the aim of developing automatic systems for ground coating and degradation analysis (Figure 2.13), urban furniture change detection (Figure 2.15(a)), urban modeling (Figure 2.15(b)), and documentation of cultural heritage (Figure 2.14).



(a) Illustrative photo.

(b) 3D point cloud.

Figure 2.13: Stop & Go Trimble TX8 acquisition to analyze ground coating and ground degradation.

2.7 Conclusions

We have presented an overview on laser scanning technologies used to acquire 3D data in urban environments. In particular, the mobile laser scanning (MLS) and the Stop & Go (S&G) mapping systems developed in the framework of TerraMobilita project have been described. We have also presented several public 3D databases in the state of the art as well as TerraMobilita databases used later in this thesis to benchmark our methods. These databases are presented as lists of XYZ Cartesian geo-referenced points with additional features such as reflectance, sensor position and ground truth annotations.

The main advantage of our annotated databases with respect to others found in the state of the art is that an *id* and a *class* are given for each 3D point. It allows the point-wise benchmarking of detection, segmentation and classification methods. In that sense, several evaluation methods have been developed in order to quantify segmentation quality and classification performance (Brédif et al., 2014). These evaluations will be presented in following chapters of this document.

In general, ground truth annotations are carried out in a manually assisted way, as it is the case of Paris-rue-Madame (Serna et al., 2014b) and TerraMobilita/iQmulus (Brédif et al., 2014) databases. In these databases, manual annotation speed was approximately 50 m/h. In spite of good annotations quality, this process is time consuming, which makes manual methods unpractical for large scale applications. In a city like Paris, with 1700 km of streets, approximately 4 years will be required for a complete manual annotation. This is one of the main motivations for the development of automatic methods for 3D semantic analysis in urban environments. The contributions of this Ph.D. thesis on this topic will be presented in following chapters of this document.

In future works, other datasets acquired in the framework of TerraMobilita project will be annotated and made available to the scientific community. For further information the reader is encouraged to visit:
http://cmm.ensmp.fr/~serna/downloads.html

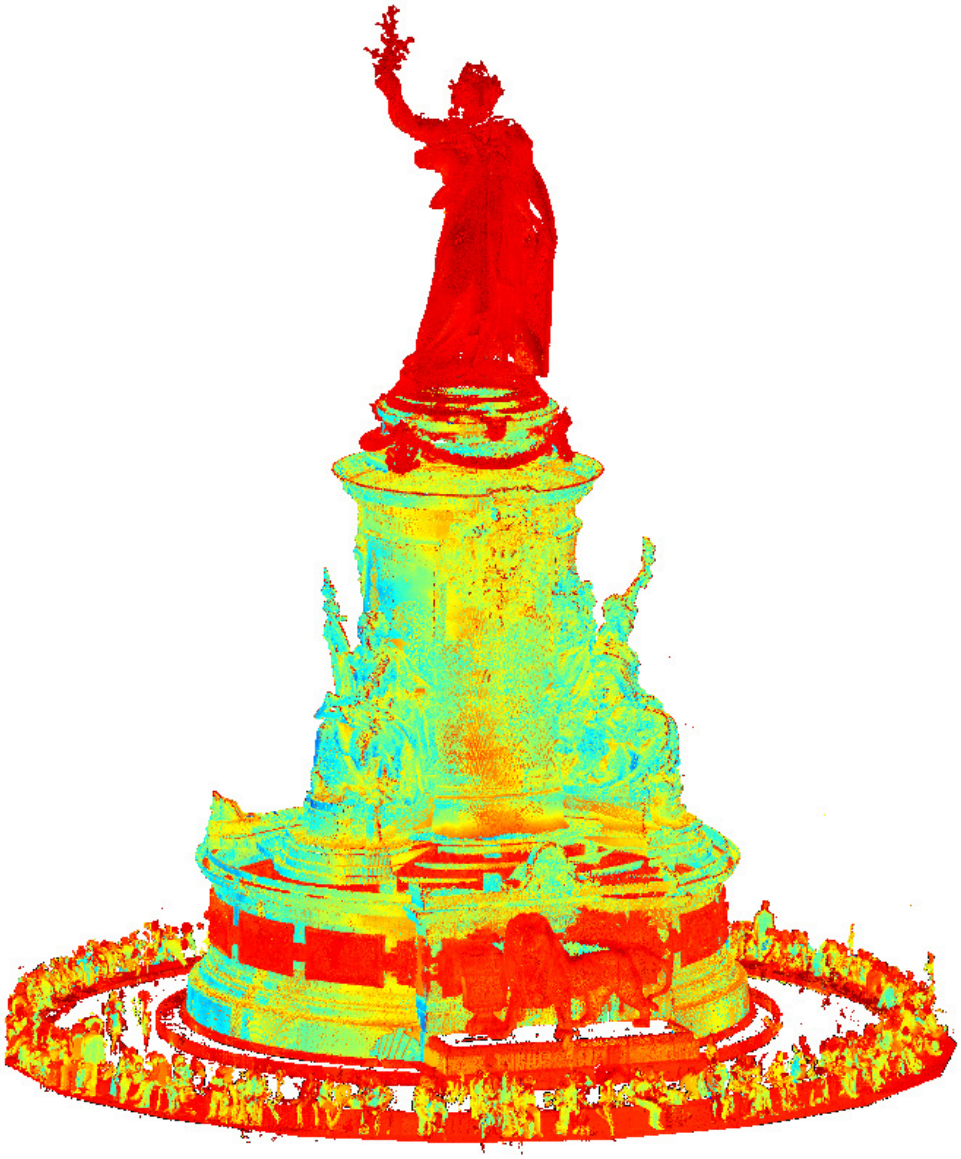
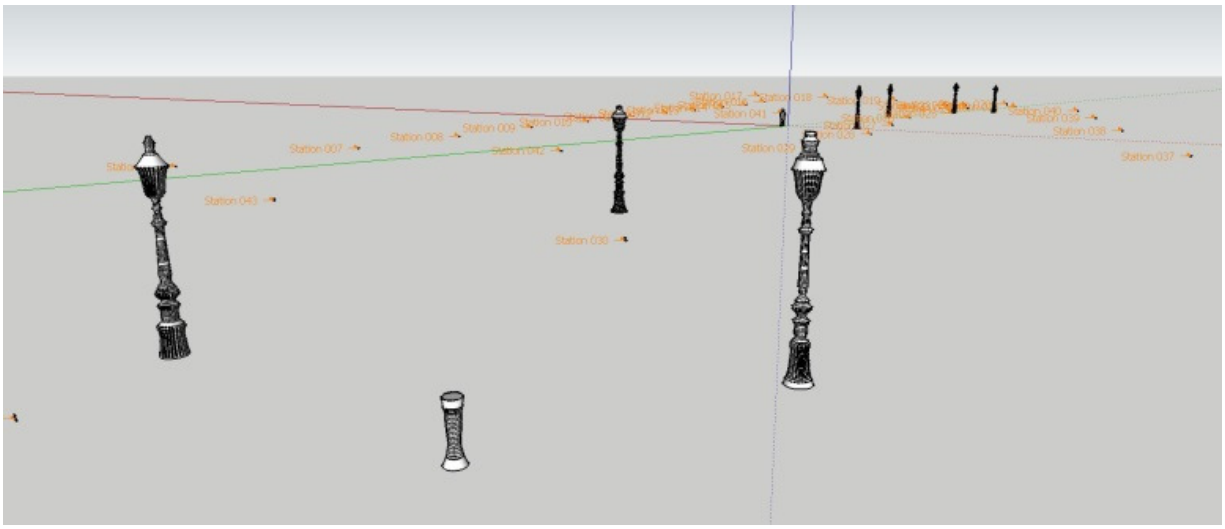


Figure 2.14: Stop & Go Trimble TX8 acquisition from *Republic square* in Paris, France.

(a) 3D point cloud.



(b) 3D modeling.

Figure 2.15: Stop & Go Trimble TX8 acquisition for urban furniture change detection and urban modeling.

3 3D data structures and preprocessing

3.1 Résumé

Dans ce chapitre, nous présenterons une révision de l'état de l'art sur les structures de données utilisées pour le traitement et la visualisation de nuages de points 3D. Ensuite, nous exposerons notre stratégie de traitement basée sur des images d'élévation ainsi que nos méthodes de pré-traitement et décomposition de la scène 3D par des images à différentes élévations.

3.2 Introduction

When working on 3D applications, there are two main approaches: the first one, called *virtual reality*, refers to the construction of artificial models of real or imaginary objects for applications in video games, graphic design, industrial design, prototyping, among others. Generally, these scenarios are built using computer aided design (CAD) software and they are the result of the developer creativity and ability; the second approach, called *virtualized reality*, refers to the representation of real world objects using information acquired by a sensor. In general, this 3D information comes from cameras and laser sensors and it is commonly presented as a 3D point cloud.

Point clouds are delivered as long lists of (x, y, z) coordinates, possibly with attributes such as intensity, color, among others. Points are usually listed in scan line order, which is not suitable for efficient processing. A suitable data structure is not only required to inspect and to visualize 3D information, but also to process it conveniently. For example in a (x, y, z) list, it is not possible to quickly determine the neighbors of a point within a given radius. Data structures such as elevation images, triangulation, meshing, octrees and k-D trees allow this kind of processing. Choosing the proper data structure is application dependent. It is possible to combine some of them to get better results in specific tasks such as visualization, filtering, segmentation and classification. In this chapter, we briefly describe some 3D data structures proposed in the literature and discuss their advantages and drawbacks. For further information, we recommend to read the book by Vosselman and Maas (2010).

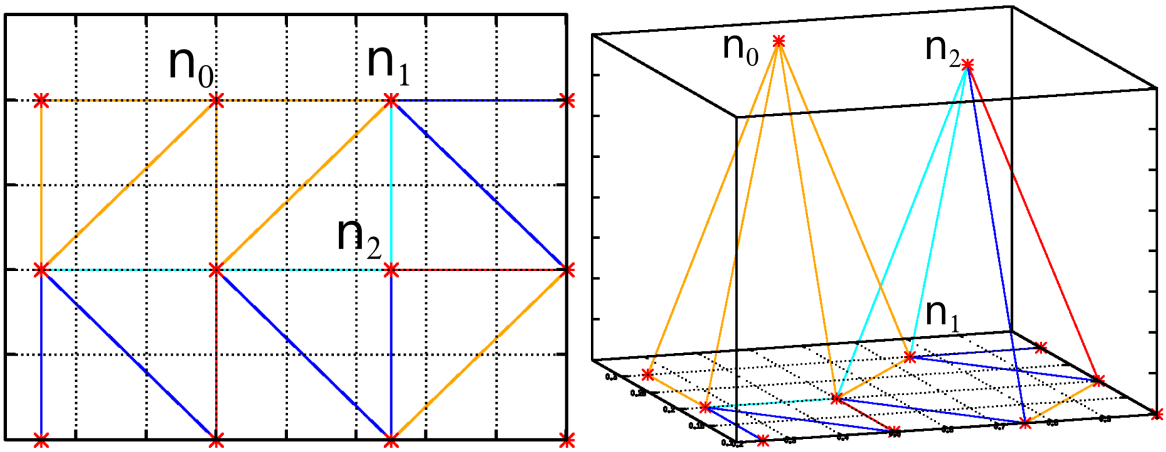
In our work, most methods work on elevation images, thus their description takes an important part in this chapter. This chapter is organized as follows: Section 3.3 describes the commonly used data structures in the state of the art. Section 3.4 explains our processing based on elevation images. Section 3.6 presents preprocessing methods used to filter and interpolate elevation images. Finally, Section 3.7 concludes this chapter.

3.3 State of the art: 3D data structures

3.3.1 Triangulation

A triangulation is a 3D meshing using triangles connecting each point in the data space. In that sense, each node corresponds to a point in the dataset. In most instances, triangles are required to meet edge-to-edge and vertex-to-vertex, and different types of triangulation may be defined depending on the object and subdivision type.

Delaunay triangulation, so named after Boris Delaunay (Delaunay, 1934), has the property that no points are inside the circumscribed circle of each triangle. Besides, it also creates compact triangles with the largest minimum angle (Okabe et al., 1992). Within a triangulated point cloud, each triangle edge defines the neighborhood relationship between points, which is useful for image processing (Jähne, 2005; Gonzalez and Woods, 2006; Serra, 1982, 1988; Soille, 2003) and computer vision (Parker, 2010) algorithms. The main drawback of this approach is that the triangulation is defined on a plane. For example, if XY plane is used, the points distribution on Z axis has no influence on the triangulation. As a consequence, points that are close in XY plane and share a triangle edge, may not be close in the 3D space, as shown in Figure 3.1. Since Delaunay triangulation is still a 2D data structure, the presence of multiple surfaces above each other may generate incorrect edges between distant points.



(a) Delaunay triangulation: 2D view

(b) Delaunay triangulation: 3D view

Figure 3.1: Example of Delaunay triangulation. Note that the triangulation is defined on a plane, thus point height has no influence on the process. As a consequence, points that are close on the plane and share a triangle edge, may not be close in the 3D space, as it is the case of points n_0, n_1 and n_2.

3.3.2 Neural networks

Neural networks (NN) can be used to model 3D objects from unstructured point clouds. In general, a training step is required and two types of models can be obtained: i) volume models, as it is the case of multi-layer feed-forward neural networks (MLFFNN). These networks are trained as classifiers in order to get binary membership functions, where a positive response is obtained for each 3D point inside the object and a negative response is obtained otherwise. ii) surface models, as it is the case of self-organizing structures. These networks provide a surface with an implicit neighborhood relation represented by the connection between near neurons. This kind of modeling can be interpreted such as a meshing and it is more appropriate than MLFFNN for processing tasks (Cretu et al., 2006).

The two most commonly used self-organizing structures are the self-organizing maps (SOM) (Kohonen, 2001; Kohonen et al., 2009) and the neural gas networks (Na et al., 2010). The input space is clustered assigning neurons to specific regions of the space. The number of inputs of each neuron is equal to the dimension of the input space. Thus, synaptic weights are interpreted as locations in this space, *i.e.* the inputs of the network are the set of (x, y, z) coordinates of the 3D point cloud. Thus, each neuron is represented by a 3D weight vector $\mathbf{m}_i \in \Re^3$. These self-organizing networks are iteratively adapted using competitive training (Martinetz et al., 1993; Fritzke, 1995). The model is asymptotically fitted to the input points according to a density function, as shown in Figure 3.2. Each region is a group of close points assigned to a neuron. After training, two input vectors belonging to the same region should be represented by one or two nearby neurons in the representation space.

Figure 3.3 shows an example of 3D meshing using SOM and NGN. The main advantage of this method is the definition of neighborhood relations, useful in post-processing tasks. Besides, NN introduces an adaptive down-sampling of the point cloud since the number of neurons is usually much lower than the number of input points. However, the main drawback is the parameter selection and the high computational cost of the training.

3.3.3 Octrees

Octrees, firstly used in 3D graphics by Meagher (1980), are 3D data structures useful for spatial indexing, streaming and data compression. Starting with a 3D cell enclosing the entire data space as the root, each internal cell containing data is recursively subdivided into up to eight non-empty octants. Each octree region uses a regular binary subdivision over all dimensions. In this way, the location of each internal cell is implicitly defined by its level and position in the octree. An octree point shares the same features with the other points in the same subdivision, and the subdivision center is at an arbitrary position inside the cell. Thus, each octree region stores additional information about the split positions. A node in an octree is similar to a node in a binary tree, with the difference that it has eight children instead of two. Additionally, each octree node contains

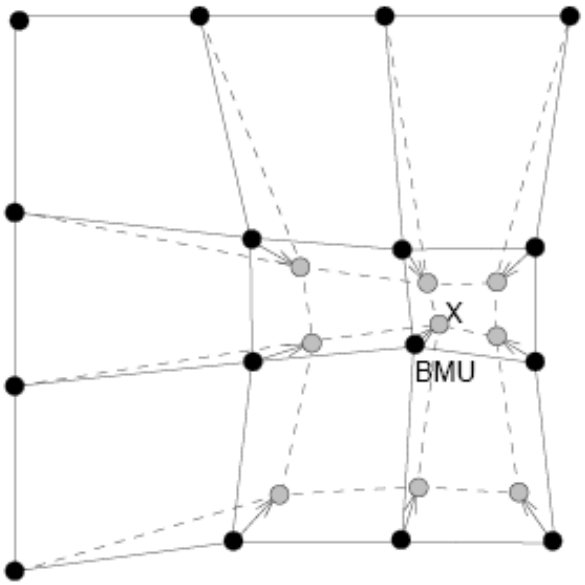
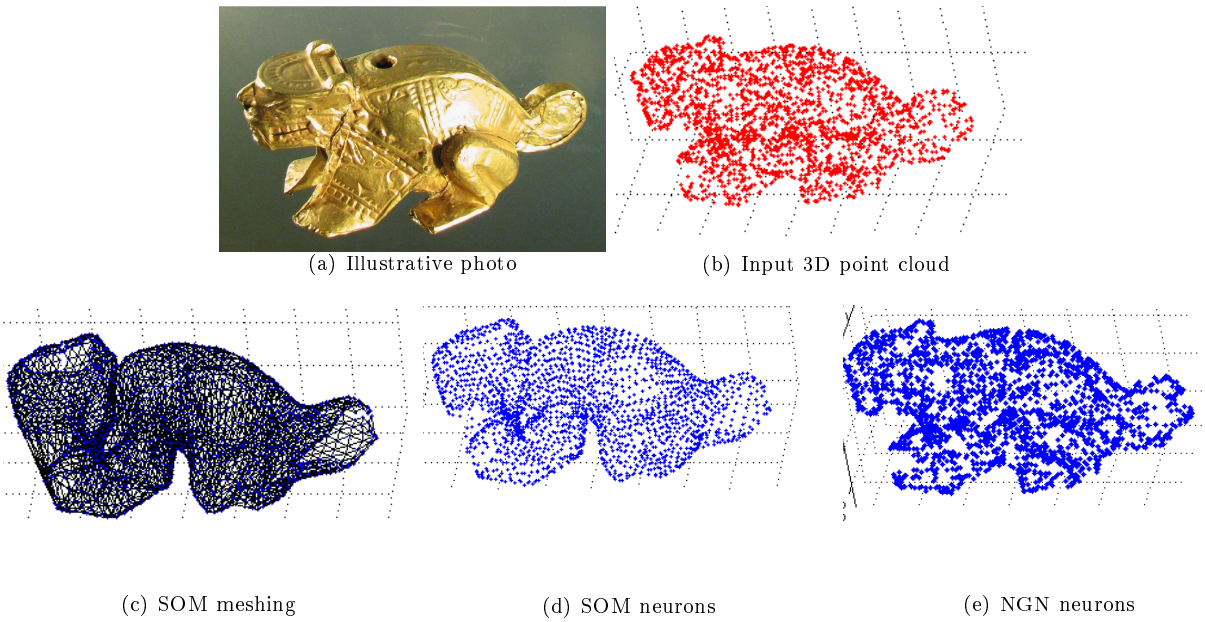


Figure 3.2: Scheme of competitive training of self-organizing neural networks. X represents a 3D point of the input point cloud, and BMU is the closest neuron to X, called the best match unit. Note that the BMU neighbors are adapted iteratively.



(a) Illustrative photo

(b) Input 3D point cloud

(c) SOM meshing

(d) SOM neurons

(e) NGN neurons

Figure 3.3: Some results of 3D modeling using self-organizing neural networks (SOM and NGN). Note that a neighborhood relation is implicitly defined by neuron connections. Images and 3D point clouds are by courtesy of Gold Museum, Bogotá, Colombia (Figueroa et al., 2006; Serna, 2011).

a key referring to their 3D coordinates. Figure 3.4 shows the recursive subdivision of a cube into octants. Note that the corresponding octree is not balanced. This is an advantage in terms of resolution since the level of detail of each branch is adapted according to the point density. However, it is a drawback in terms of computational cost because searching the neighbors of a given point depends on its location and the number of subdivisions on its branch.

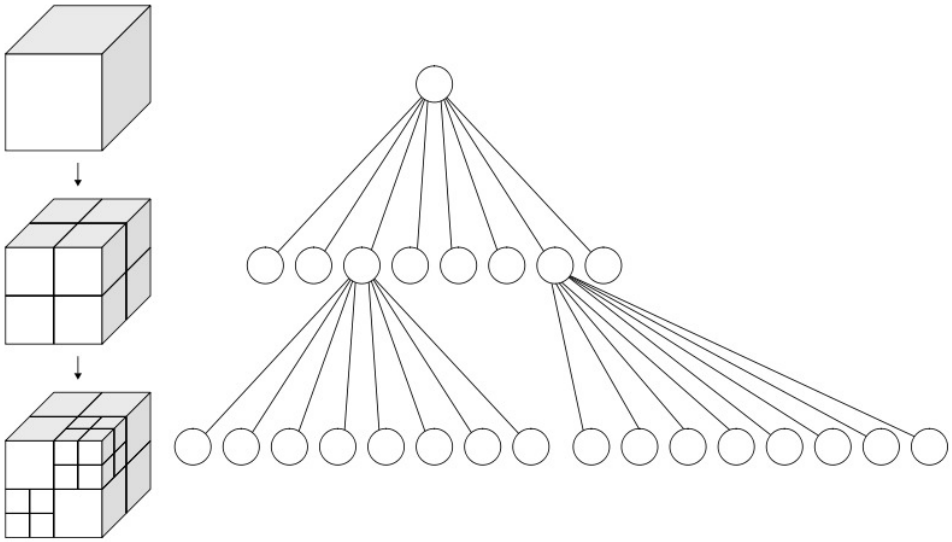


Figure 3.4: Octree: recursive subdivision of a cube into octants. Octrees have been firstly used in 3D graphics by Meagher (1980) and they are 3D data structures useful for spatial indexing, streaming and data compression. Image taken from Wikipedia, the free encyclopedia, 2012.

3.3.4 kd-trees

In contrast to octrees, k-D trees guarantee a fully balanced hierarchical data structure for datasets sampled from a k-dimensional manifold (Bentley, 1975). A k-D tree is a k-dimensional binary tree useful for spatial indexing, streaming and data compression. Each node contains pointers to the left and right subtrees. All non-leaf nodes have an orthogonal plane that divides one of the dimensions. Points in the left subtree are below this plane and points in the right subtree are above. Each leaf node contains a pointer to the list of points located in the corresponding cell, defined by the intersection of half-spaces given by nodes in the part of the root node to the leaf node itself. Using this data structure it is possible to locate a point in a k-D tree with N points in an average time of $O\{log(N)\}$. Figure 3.5 illustrates the recursive division of a cube into eight leaf cells using a 3D-tree.

3.3.5 Projection to images

In the literature, several methods project 3D information onto a 2D grid in order to reduce the problem complexity and to speed up the computational processing. As each pixel of the projected grid contains depth information, it is called range image or depth map. This kind of 2.5D image has a long tradition in the scientific community (Hoover et al., 1996) and it is of great interest nowadays due to technological developments in remote sensing equipments such as Riegl, Velodyne and Kinect sensors. Figure 3.6 presents an example of a range image acquired by Microsoft Kinect sensor in an indoor environment. It is noteworthy that the depth information is coded in the gray-scale values: the brighter the pixel, the more distant the point.

In general, different types of projections can be defined. Gorte (2007) projects TLS data to a plane from the sensor point of view. As a result, a "panoramic" range image is obtained. In a similar way, Zhu et al. (2010) generate range images in which rows represent the acquisition time of each laser scan–line, columns represent the sequential order of measurement and pixel values code the distance from the sensor to the point. Another approach consists in projecting the 3D point cloud to an spheric coordinate system with origin in the laser location. This solution is called "station view" and is commonly used in commercial solutions such as Trimble RealWorks (Trimble, 2014b).

When the range image plane corresponds to the ground plane (i.e. the horizontal plane), the depth information is related with the height of the objects in the scene. Such images represent a nadir view of the scene and are commonly called elevation images in the remote sensing community. This is the projection used by us and by other several authors in the state of the art (Hernández and Marcotegui, 2009c; Golovinskiy et al., 2009; Weinmann et al., 2013). This approach is specially adapted to segment ground and urban objects.

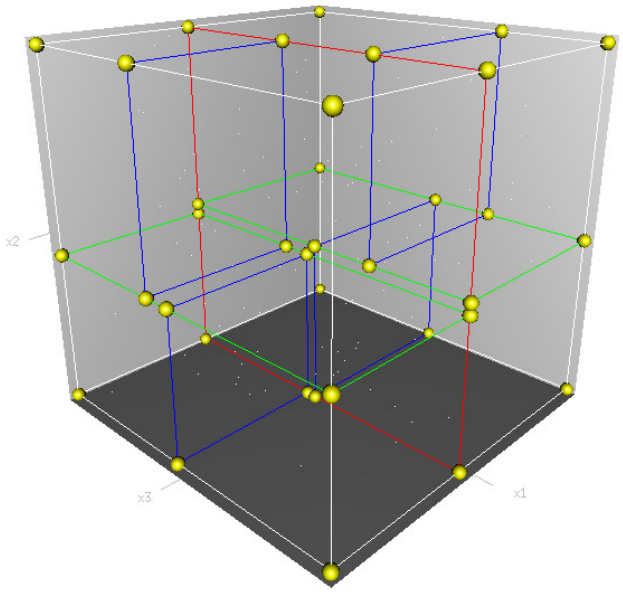


Figure 3.5: Example of a 3D tree. The first split (red) cuts the root cell (white) into two sub-cells, each of which is then split (green) into two sub-cells. Finally, each of those four is split (blue) into two sub-cells. Since there is no more splitting, the final eight are called leaf cells. Image taken from Wikipedia, the free encyclopedia, 2012.

Figure 3.6: Example of Kinect range image. In this case, the range image is the direct output of the Kinect sensor, thus no additional projection is required. Courtesy of Vincent Weistroffer, Robotics Laboratory (CAOR), MINES ParisTech, 2012.

3.3.6 Discussion

According to the state of the art, several data structures are available in order to visualize and process 3D data. Several methods perform directly on the 3D point cloud, on a point by point basis. In general, these methods are the most expensive in computational terms (Mallet et al., 2008; Demantke et al., 2010; Rutzinger et al., 2011; Pu et al., 2011). Slightly less expensive methods process the 3D point cloud on a strip by strip basis

(Owechko et al., 2010; Vosselman and Zhou, 2009; Zhou and Vosselman, 2012). The main drawback is that intrinsic information between neighboring strips is missing. To solve this problem, a neighborhood relation can be defined using 3D data structures such as meshing (Cretu et al., 2006; Schnabel et al., 2008; Serna, 2011), voxelization (Douillard et al., 2011), fitting primitives (Owechko et al., 2010; Poreba and Goulette, 2012b,a) and projection images (Gorte, 2007; Kammel et al., 2008; Ferguson et al., 2008; Hernández, 2009; Zhu et al., 2010; Serna and Marcotegui, 2014). The selection of the appropriate data structure is application dependent and it should be a trade-off between performance and computational cost.

In this work, elevation images are the 3D data structure on which our methods are based on. 3D point clouds are projected to elevation images because they are convenient structures to visualize and to process data. One can utilize all the large collection of existing image processing tools, in particular mathematical morphology (Serra, 1982; Soille, 2003). Additionally, images can be processed quickly, implicitly define neighborhood relationships and require less memory than 3D data.

The rest of this chapter is devoted to the image elevation generation and the preprocessing methods used to filter and interpolate elevation images.

3.4 3D processing using elevation images

In general, the idea of deriving elevation images from 3D point clouds is not new. In previous works (Gorte, 2007; Kammel et al., 2008; Ferguson et al., 2008; Zhu et al., 2010; Niemeyer et al., 2014), the authors used elevation images to process urban 3D data. In particular, the works by Hernández and Marcotegui (2009c) and Hernández (2009) have been the starting point of this chapter. Their projection and interpolation methods (Sections 3.4.1 and 3.6.3) have been applied identically as in these works. Nevertheless, the projection by slices and the filtering methods are entirely original and constitute one of the main contributions of this chapter. Several contributions of this chapter have already been published in Serna and Marcotegui (2013b).

3.4.1 Elevation image generation

Elevation images are 2.5D structures that contain height information at each pixel. They are generated by an orthographic projection to a virtual camera plane, *i.e.* the height is the distance from each 3D point to the projection plane. In general, the camera plane is chosen to coincide with the horizontal plane. The camera model $\mathcal{P}$ is a projective transformation from $\mathbf{R}^3 \rightarrow \mathbf{N}^2$, and it can be decomposed in three sequential transformations as follows:

Definition 3.4.1 *Let $M = (X, Y, Z)$ be a 3D point in $\mathbf{R}^3$ and $m = (u, v)$ a point in the image space $\mathbf{N}^2$. The camera model $\mathcal{P}$ is defined as the successive transformations:*

$$(X, Y, Z) \xrightarrow{T} (X_c, Y_c, Z_c) \xrightarrow{P} (x, y) \xrightarrow{A} (u, v) \tag{3.1}$$

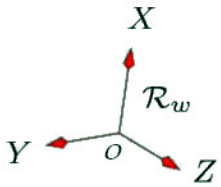
$$[T] = \begin{bmatrix} [Rot] & t \\ 0^T & 1 \end{bmatrix}$$

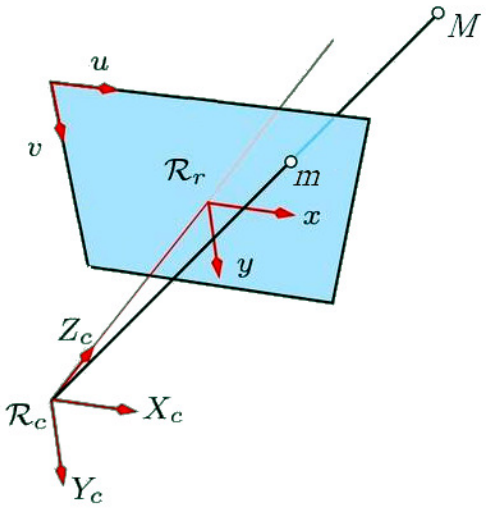
Rot: Rotation matrix
t: translation vector

$$[P] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & foc^{-1} & 0 \end{bmatrix}$$

foc: focal length

$$[A] = \begin{bmatrix} k_x & 0 & c_x \\ 0 & k_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$





where, (c_x, c_y) is the intersection point between the optical axis and the image plane, (k_x, k_y) are the number of pixels per length unit, $\mathcal{R}_w$ is the real world coordinate system, $\mathcal{R}_c$ is the camera coordinate system with origin in the optical center, and $\mathcal{R}_r$ is the projection plane coordinate system. The virtual camera is chosen such that:

- *The camera plane corresponds to the horizontal plane with normal vector* $\overrightarrow{n} = (0, 0, 1)$ *and crossing the lowest point in the point cloud* $(0, 0, z_{min})$.
- *Rotation matrix* $[Rot]$ *is equal to the identity.*
- *The point cloud and the projection plane are centered on the gravity center of the point cloud. Thus, translation vector* t *is equal to the gravity center, and the intersection point* (c_x, c_y)=(0,0).
- *The projection is orthographic. Thus, the projection axis is orthogonal to the projection plane and the projection center is located at the infinity. It means,* $foc = \infty$, $x = X_c$ *and* $y = Y_c$.

According to these assumptions, the number of pixels per length unit (k_x, k_y) are the only free parameters. They have to be carefully chosen in order to avoid connectivity problems and loss of information. Additionally, this selection should imply a trade off between performance and computational cost. This parameter is discussed later in Section 3.4.2.

During projection, several points can be projected on the same pixel. Thus, four images are defined:

- *Maximal elevation image* (f), or simply elevation image, stores the maximal elevation (vertical distance from each 3D point to the projection plane) among all projected points on the same pixel. This image contains the surface of the scene from a nadir view.
- *Minimal elevation image* (f_{min}), stores the minimal elevation among all projected points on the same pixel. This image generally contains the lowest objects and the ground.
- *Relative height image* $(f_{\text{height}} = f - f_{min})$, contains the difference between maximal and minimal elevation images. This image allows to estimate the height of objects such as facades or lampposts, independently from the street slope. However, note that this image does not contain the complete object height information because the ground is not always visible.
- *Accumulation image* (f_{acc}), stores the number of points projected on each pixel. This image is very useful to detect vertical high structures in the scene. However, a normalization is required as explained later in Section 3.6.

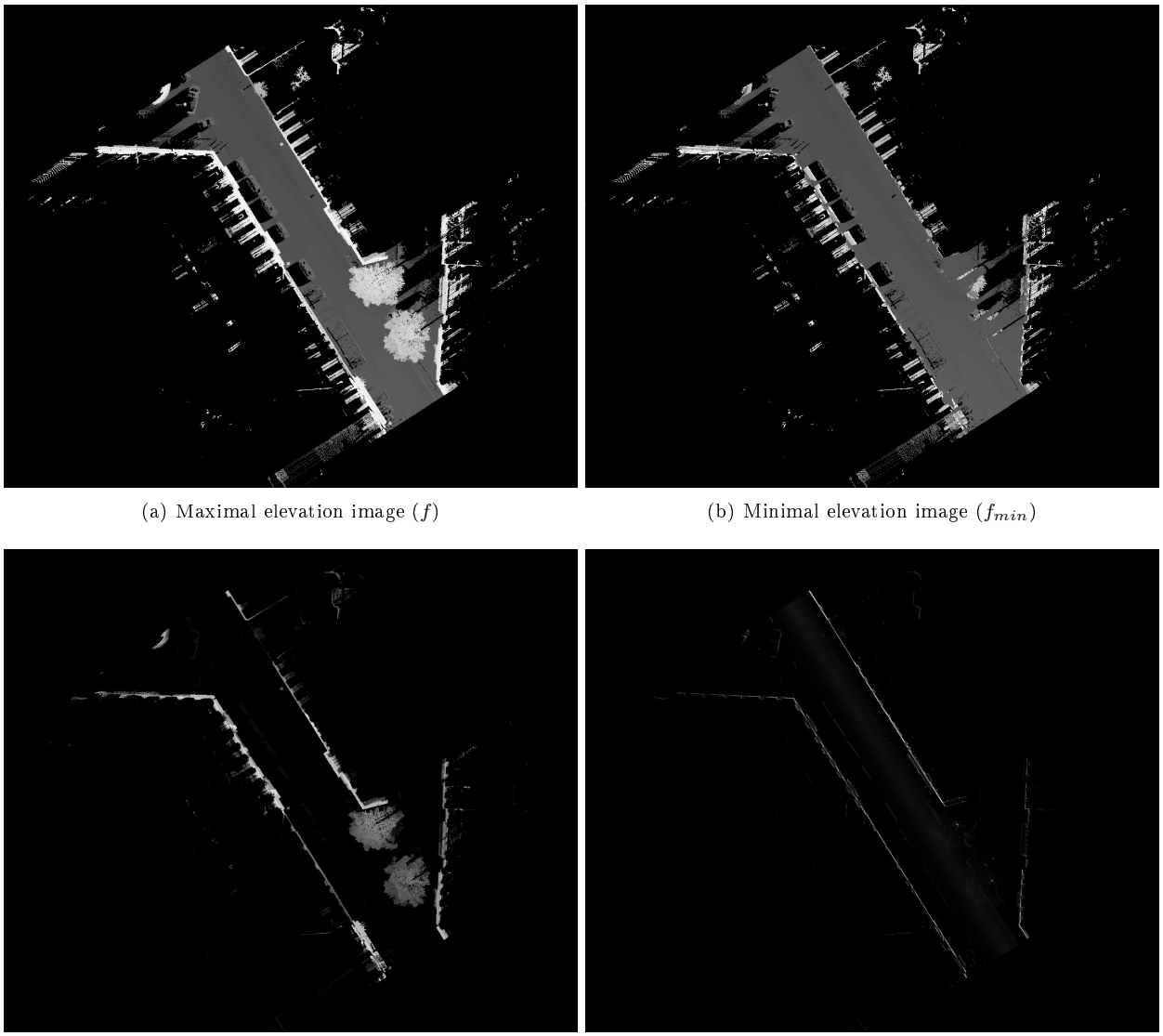
Figure 3.7 shows an example of these four images in a test site in *rue d'Assas* in Paris. On the one hand, Figure 3.7(a) presents the maximal elevation image, note that the maximal distance is stored for each pixel. This corresponds to a nadir view of the urban scene. On the other hand, Figure 3.7(b) presents the minimal elevation image, where the minimal distance is stored for each pixel. Note that this image is particularly appropriate for analysis at the ground level since high objects such as trees do not appear on the image. Actually, this image is used for the accessibility analysis presented later in Chapter 4. Figure 3.7(c) presents the relative height image f_{height} computed as the difference between the maximal and minimal elevation images. Note that relative height on several points such as those on the car roofs is zero, because the ground is occluded and the maximal and minimal elevated points are the same. Figure 3.7(d) presents the accumulation image, defined as the number of points projected on the same pixel. Note that these two latter images (f_{height} and f_{acc}) are very useful to detect vertical high structures such as facades and pole-like objects. This segmentation problem will be discussed later in Chapter 6.

In general, our processing steps are performed on images f and f_{min} while other two images f_{height} and f_{acc} are used to support decisions during the analysis or to compute object features. Figure 3.8 presents two 3D point clouds and their corresponding elevation images for a test site in Paris, France.

Elevation images imply a reduction in the amount of data to be processed with respect to the 3D point cloud. Moreover, neighborhood relationships are given on the image without any additional computing. In other words, we process an elevation image using image processing techniques, which is much faster than processing the 3D points directly. Additionally, at the end of the process, the semantic analysis results can be reprojected onto the 3D point cloud if the results should be visualized in 3D. An analysis on the elevation image size with respect to k parameter can be found in the following subsection.

3.4.2 Elevation image resolution

According to the assumptions made in Definition 3.4.1, the number of pixels per length unit (k_x, k_y) are the only free parameters during the elevation image generation. In order to simplify this selection, we consider square pixels assuming $k_x = k_y = k$, where k has to be carefully chosen. On the one hand, if k is too small, fine details



(a) Maximal elevation image (f)

(b) Minimal elevation image (f_{min})

(c) Relative height image ($f_{\text{height}} = f - f_{min}$)

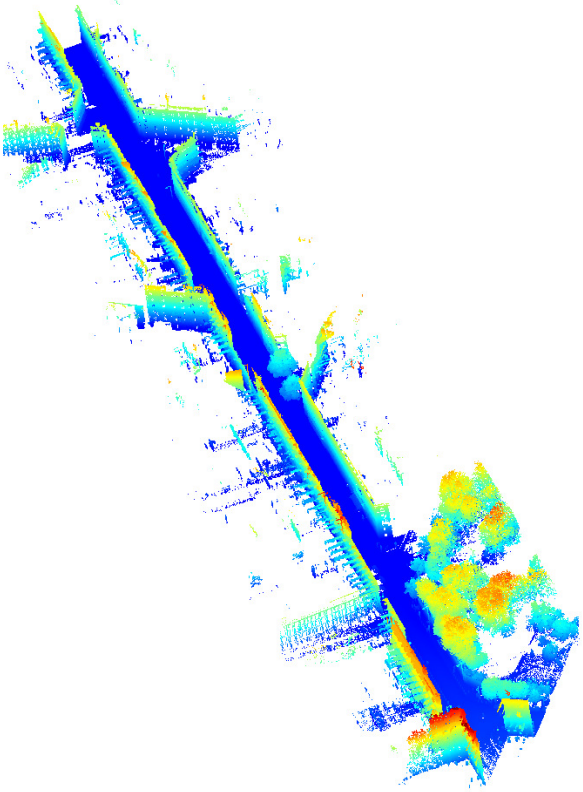
(d) Accumulation image (f_{acc})

Figure 3.7: 3D point cloud and elevation images for a test site in *rue d'Assas* in Paris, France. Data acquired by Stereopolis II, IGN©France. (a) maximal elevation image, containing the maximal distance for each pixel. This corresponds to a nadir view of the urban scene. (b) minimal elevation image, where the minimal distance is stored for each pixel. Note that this image is particularly appropriate for analysis at the ground level since high objects such as trees do not appear on the image. (c) relative height image, computed as the difference between the maximal and minimal elevation images. (d) accumulation image, defined as the number of points projected on the same pixel.

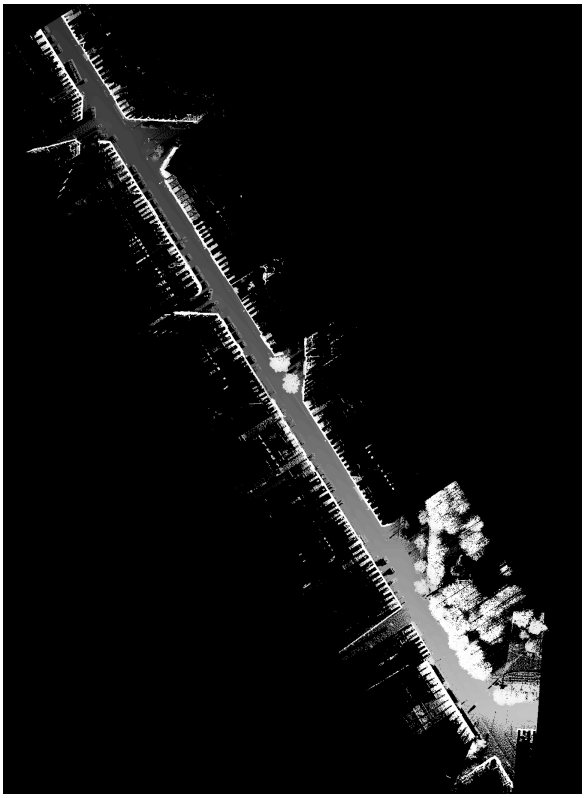
are not preserved because too many points would be projected on the same pixel. On the other hand, too large k presuppose connectivity problems and large image sizes, which implies high computational time that would no longer justify the use of elevation images instead of 3D point clouds. This parameter is the most critical in terms of quality vs. processing time since it determines the elevation image resolution. Let us introduce the problem of the processing time taking the image size into account.

Consider the two datasets of Figure 3.8, in the 6^{th} Parisian district. They contain MLS data from approximately 500 m of *rue d'Assas* and 300 m of *rue Cassette* in Paris, respectively. These data have been acquired by Stereopolis II, a MLS system from IGN France (Paparoditis et al., 2012). Table 3.1 presents some technical specifications.

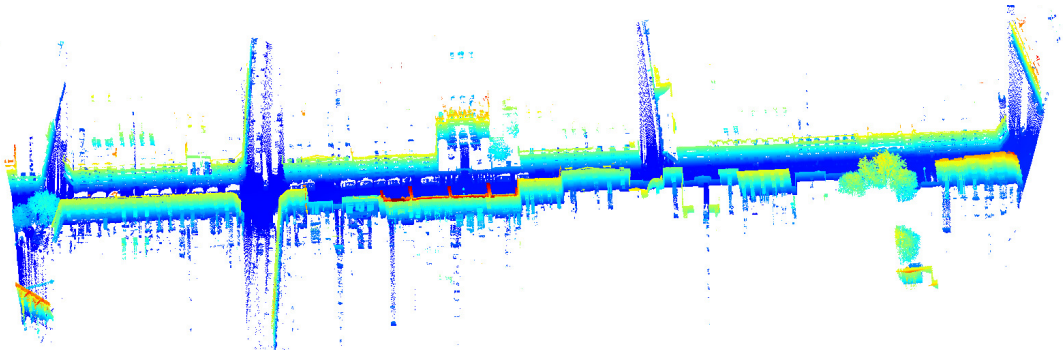
The elevation image size has a critical effect in the computation time: the bigger the image, the slower the



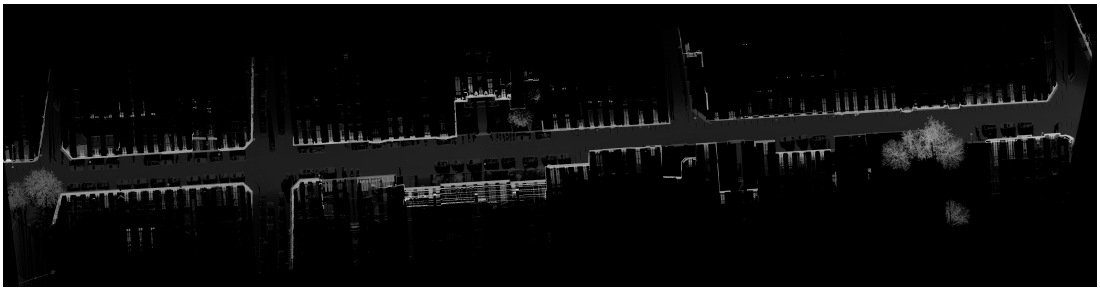
(a) Rue d'Assas: 3D point cloud



(b) Rue d'Assas: Elevation image



(c) Rue Cassette: 3D point cloud



(d) Rue Cassette: Elevation image

Figure 3.8: 3D point cloud and elevation image for two test sites in *rue d'Assas* and *rue Cassette* in Paris, France. Data acquired by Stereopolis II, IGN©France.

Table 3.1: TerraMobilita datasets from *rue d'Assas* and *rue Cassette* in Paris. IGN©France.

	rue d'Assas	**rue Cassette**
Street length	500 m	300 m
3D points	24×10^6 points	18×10^6 points
Acquisition time	2'25"	1'31"
Vehicle speed (average)	12.4 km/h	11.9 km/h

computation. Figure 3.9 shows the elevation image size for different k values from 5 pix/m (20×20 cm^2/pix) to 20 pix/m (5×5 cm^2/pix). It is noteworthy that the number of pixels increases as k^2.

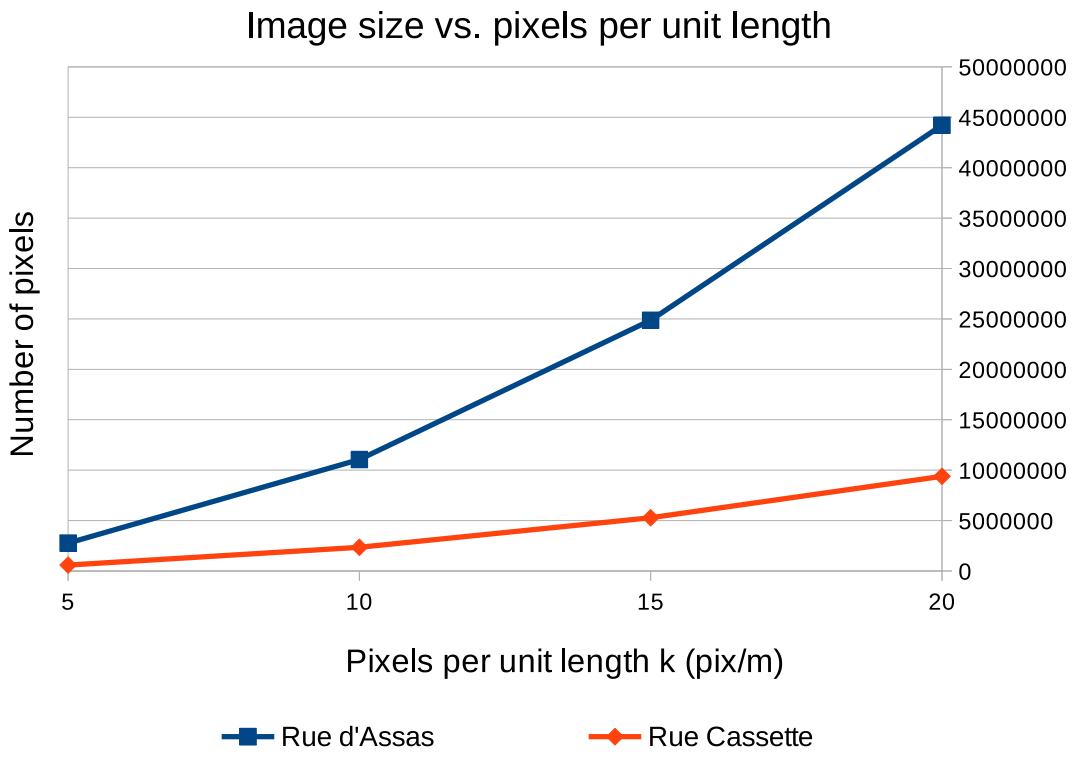


Figure 3.9: Elevation image size for different k parameters (number of pixels per length unit) from 5 pix/m to 20 pix/m. The elevation image size has a critical effect in the computation time: the bigger the image, the slower the computation. It is noteworthy that the number of pixels increases as k^2.

In *rue d'Assas* site (Figure 3.8(a)), elevation image sizes are 1390×1988 pixels for k=5 pix/m and 5560×7950 pixels for k=20 pix/m. Note that its elevation image (Figure 3.8(b)) has many black pixels (no-data points) due to the diagonal street direction (South East-North Western). Only 27% of the pixels contains information. Applying an appropriate rotation to the input point cloud can reduce the number of black pixels in the image projection. The aim is matching the principal axis of the point cloud with the X-axis (or Y-axis) of the elevation image. Another solution consists in defining a mask in order to ignore black pixels during the processing.

In *rue Cassette* site (Figure 3.8(c)), elevation image sizes are 389×1508 for k=5 pix/m and 1556×6031 for k=20 pix/m. In this case, the elevation image (Figure 3.8(d)) is better distributed due to the vertical street direction (South-North). In this case, 50% of the points contains information.

Processing for k=5 pix/m provides the fastest results, while processing for k=20 pix/m provides the most accurate ones. In the case of large scale applications, strict time constraints are not required. However, a trade-off between speed and accuracy is desired. As it will be demonstrated in the following chapters, our methods

are fast since their process time is comparable to the one required for the acquisition. One of the advantages of using elevation images is that 3D points are projected to an image and they are processed as a complete set using digital image processing techniques, preserving intrinsic neighboring information. Details about our processing methods will be discussed in the following chapters.

3.5 Elevation images by slices

One of the disadvantages processing 3D urban data using elevation images is that high objects may occlude lower objects located under them. For example, in Figure 3.10, the pedestrian in the right part (object ⑤) does not appear on an elevation image because it is below a tree (object ⑥). This can be appreciated in a real scenario of Figure 3.14 in *St. Sulpice* square, where several cars and pedestrians do not appear in the elevation images due to high trees. To solve this problem, we propose a projection strategy using slices. Particularly, two slices are used: i) a lower slice, containing points between the ground level and a given height H_{slice} in the vertical axis. This slice is built to contain most of urban objects; and, ii) an upper slice, containing points higher than H_{slice}. This slice contains the highest objects such as facades, treetops, lampposts and off-ground objects.

In our experiments, H_{slice} has been experimentally set to 3.5 m, *i.e.* obstacles for a walking pedestrian, as marked by the blue dotted line in Figure 3.10. This threshold can be modified in order to define obstacles maps at different heights according to different types of mobility: children, persons using a wheelchair, etc. Note that in the case of a non-flat surface, it is very important to segment the ground in order to adapt each slice to the ground curvature. Methods used to segment the ground will be discussed later in Chapter 4.

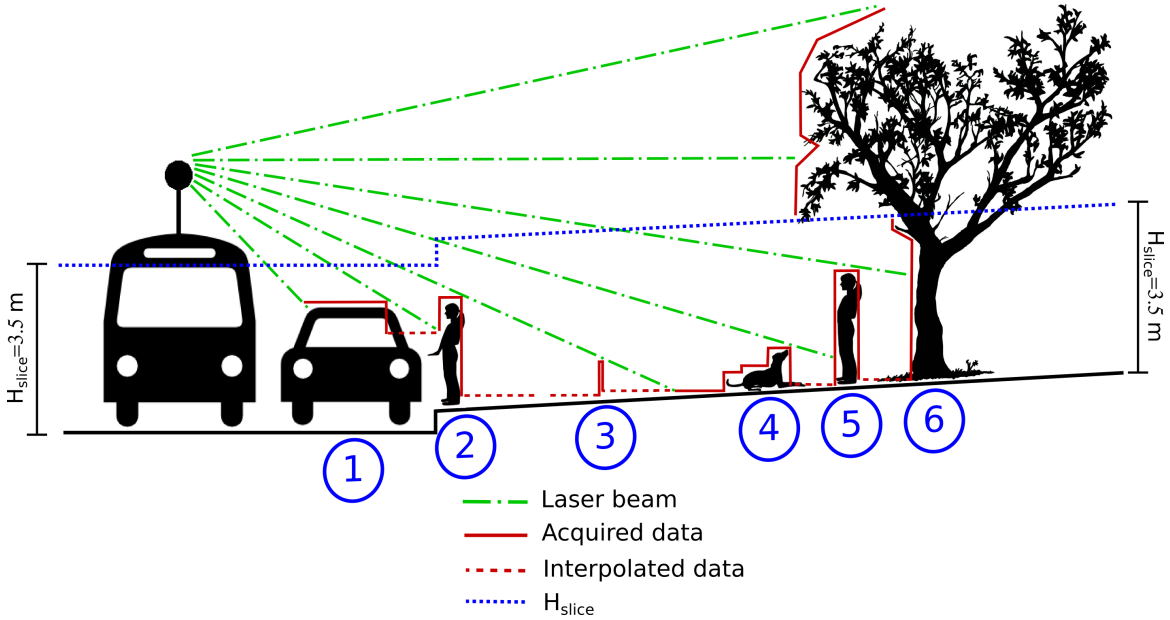
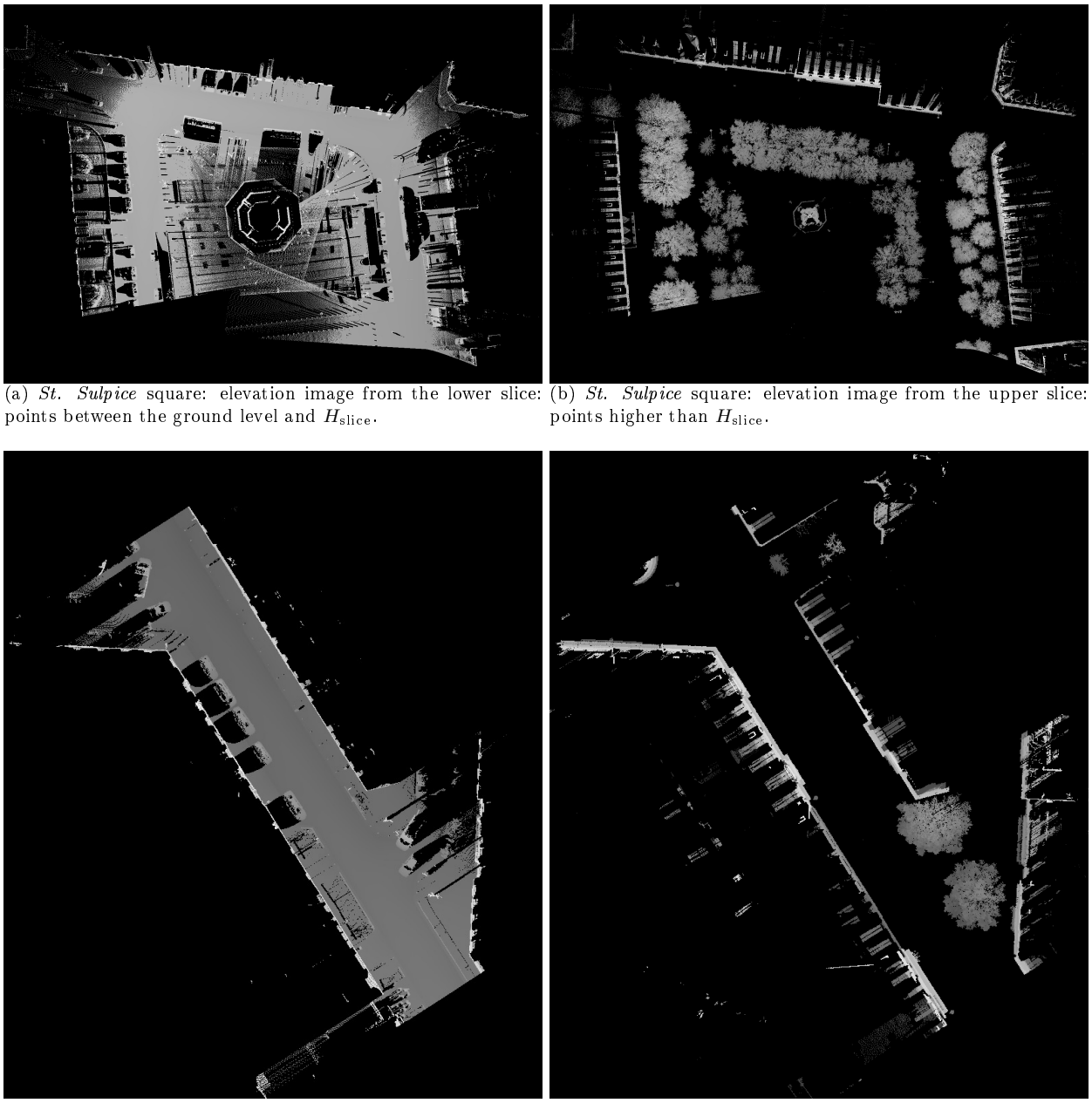


Figure 3.10: Projection by slices on the 1D case. The urban scene contains the following urban objects enumerated from ① to ⑥: ① car, ② pedestrian, ③ noise, ④ dog, ⑤ pedestrian and ⑥ tree. Note that processing by slices is useful to avoid that high objects such as trees (object ⑥) occlude lower objects such as pedestrians (object ⑤).

Figure 3.11 shows an example of two slices in two experimental sites in the 6^{th} Parisian district. It is noteworthy that this processing based on slices is particularly adapted to urban environments.

This processing based on slices can be interpreted as an adaptive voxelization using surfaces parallel to the ground. On the one hand, voxel dimensions are determined by the image pixel size and the slice height. On the other hand, voxel position is determined by the ground. One of the main advantages is that each slice can be analyzed using image processing techniques not only in the 2D space, *i.e.* slice by slice, but also in the 3D space, *i.e.* keeping 3D neighborhood relation between neighboring slices. It is noteworthy that in the case of high objects such as trees, a part of the objects is in the lower slice while the other part is in the upper slice,



(a) *St. Sulpice* square: elevation image from the lower slice: points between the ground level and H_{slice}.

(b) *St. Sulpice* square: elevation image from the upper slice: points higher than H_{slice}.

(c) Rue d'Assas: lower slice elevation image: points between the ground level and H_{slice}.

(d) Rue d'Assas: upper slice elevation image: points higher than H_{slice}.

Figure 3.11: Generating elevation images by slices for two test sites in *St. Sulpice* square and *rue d'Assas* in Paris, France. Data acquired by Stereopolis II, IGN©France. In our experiments, H_{slice} has been experimentally set to 3.5 m, *i.e.* obstacles for a walking pedestrian, as marked by the blue dotted line in Figure 3.10. This threshold can be modified in order to define obstacles maps at different heights according to different types of mobility: children, persons using a wheelchair, etc.

neighborhood relation between slices is very important in order to retrieve objects separate during individual processing of each slice. In the case of more detailed analysis, several slices may be defined.

3.6 Image preprocessing

Before our semantic analysis, several preprocessing techniques to filter and to interpolate elevation images are applied. These techniques are presented in this chapter in order to avoid repetition each time they are used in the rest of this manuscript.

3.6.1 Filtering distant points

In the 3D point cloud, too distant points have two main drawbacks during the processing:

- The first one, common to all methods, is that they are not reliable since the sensor precision decreases as the distance. In general, data precision depends on the technical specifications of the acquisition system, the distance from the sensor to the object, the object material and color, the angle of incidence of the beam, among others. In this preprocessing step, we are interested in the distance from the sensor to the object: the bigger the distance, the lower the precision. Thus, it is desirable to filter these distant points.
- The second one, inherent to our processing strategy, is that too distant points may produce enormous projection images with big empty zones. In order to speed up our processing, these isolated points should be eliminated. Moreover, these points do not contain any useful information.

Using the trajectory information of the acquisition system, it is possible to compute the distance from each 3D point P_i to the sensor at the acquisition moment. It is called the *radius*, and it is computed simply as the 3D Euclidean distance $r_i = \sqrt{(x_i - xs_i)^2 + (y_i - ys_i)^2 + (z_i - zs_i)^2}$, where x_i, y_i and z_i stand for the 3D coordinates of the input point P_i, and xs_i, ys_i and zs_i stand for the sensor position at the acquisition moment. In some datasets, this computation is not needed since the *radius* is given directly in the point cloud file, as it is the case of Stereopolis II system (Paparoditis et al., 2012).

In order to filter out distant points, a simple threshold is applied to the *radius* of each point. According to technical and practical issues explained before in Chapter 2, *i.e.* the maximum range of the laser, points farther than 50 m are not considered. Figure 3.12 shows the effect of filtering distant points in a 3D point cloud from *rue d'Assas* in Paris. Note that the elevation image computed from the filtered point cloud (Figure 3.12(b)) contains approximately 30 times less pixels than that computed from the original point cloud (Figure 3.12(a)). Analyzing the histogram of Figure 3.12(c) and Figure 3.12(d), we can see that relevant information is kept in the filtered point cloud.

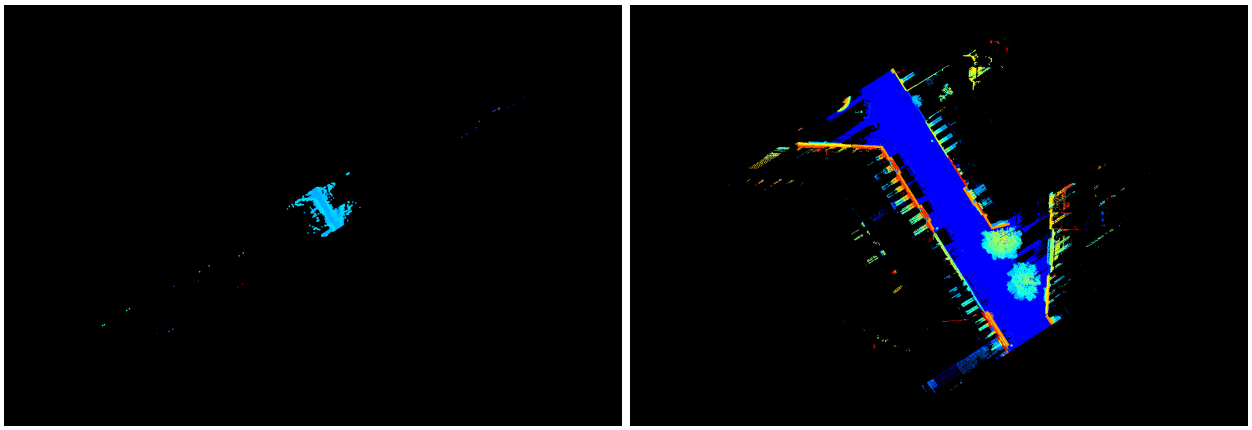
3.6.2 Filtering redundant points

When working with dense data, redundant information should be properly managed in order to get reliable results and to reduce processing time. The point cloud density depends on the sensor technical specifications, the sensor-to-object distance and the vehicle trajectory/speed during the acquisition. With respect to the first two cases, Riegl and Velodyne sensors have been used in our experiments and their technical specifications have been summarized in Chapter 2. In the third case, when the vehicle goes slowly or when it stops, *e.g.* due to traffic lights or traffic jams, the point density increases since scan lines overlap, giving great values on the accumulation image.

In order to reduce redundant data, we propose a filtering strategy taking advantage of the way each acquisition profile is acquired. As aforementioned in Chapter 2, each spin of the laser sensor represents an acquisition profile. Giving an identifier to each individual profile, it is possible to determine overlapping profiles. When two or more acquisition profiles overlap, only the profile providing the maximal information is kept. This filtering is carried out pixel by pixel on the accumulation image, thus the result is a combination of segments from individual profiles.

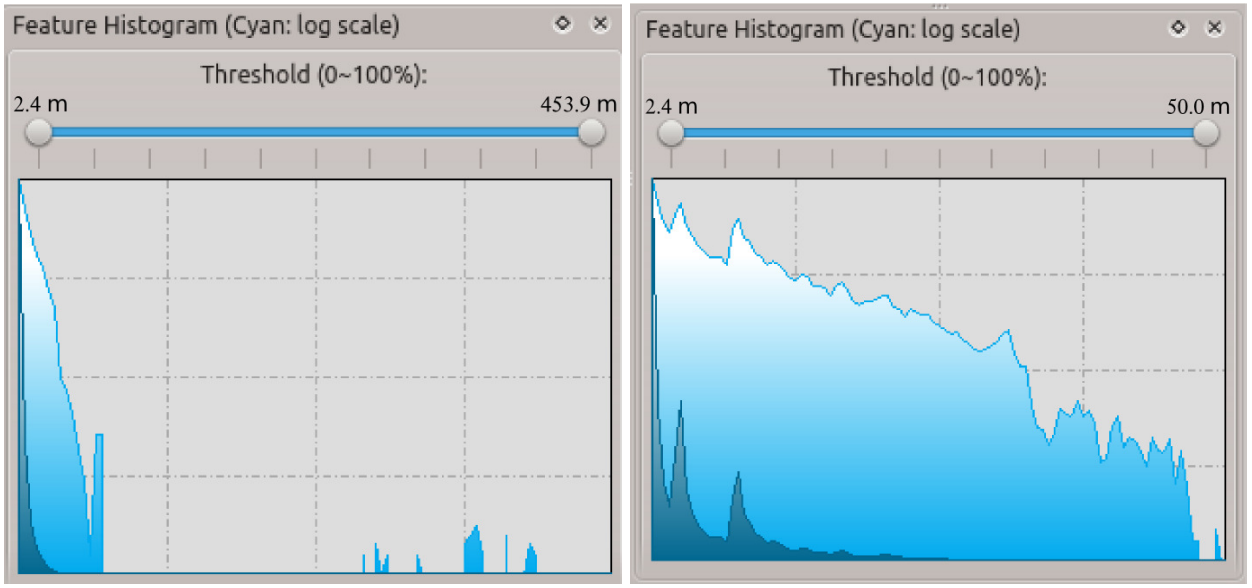
Let us explain this filtering strategy through the example of Figure 3.13. Consider the acquisition vehicle stopped during the interval $[t_0, t_2]$ due to a traffic light and the three overlapping scan lines $f_{t=t_0}(x)$, $f_{t=t_1}(x)$ and $f_{t=t_2}(x)$, as shown in Figure 3.13(a), Figure 3.13(b) and Figure 3.13(c). The urban scene contains the following urban objects enumerated from ① to ④: ① bollard, ② pedestrian, ③ pedestrian and ④ house facade.

On the one hand, let us analyze the two fix objects ① and ④. These objects appear in the three profiles, then their accumulation f_{acc} is three times the accumulation of an individual profile, as shown in Figure 3.13(d). On the other hand, let us analyze the two mobile objects ② and ③. These objects only appear in some profiles, then their accumulation f_{acc} not only depends on the object geometry but also on the time spent by the pedestrians



(a) Elevation image from the original point cloud. Image size: 7523 × 4206 = 31.6×10^6 pixels

(b) Elevation image after filtering distant points in the input point cloud. Image size: 1131 × 952 pixels = 1.1×10^6 pixels



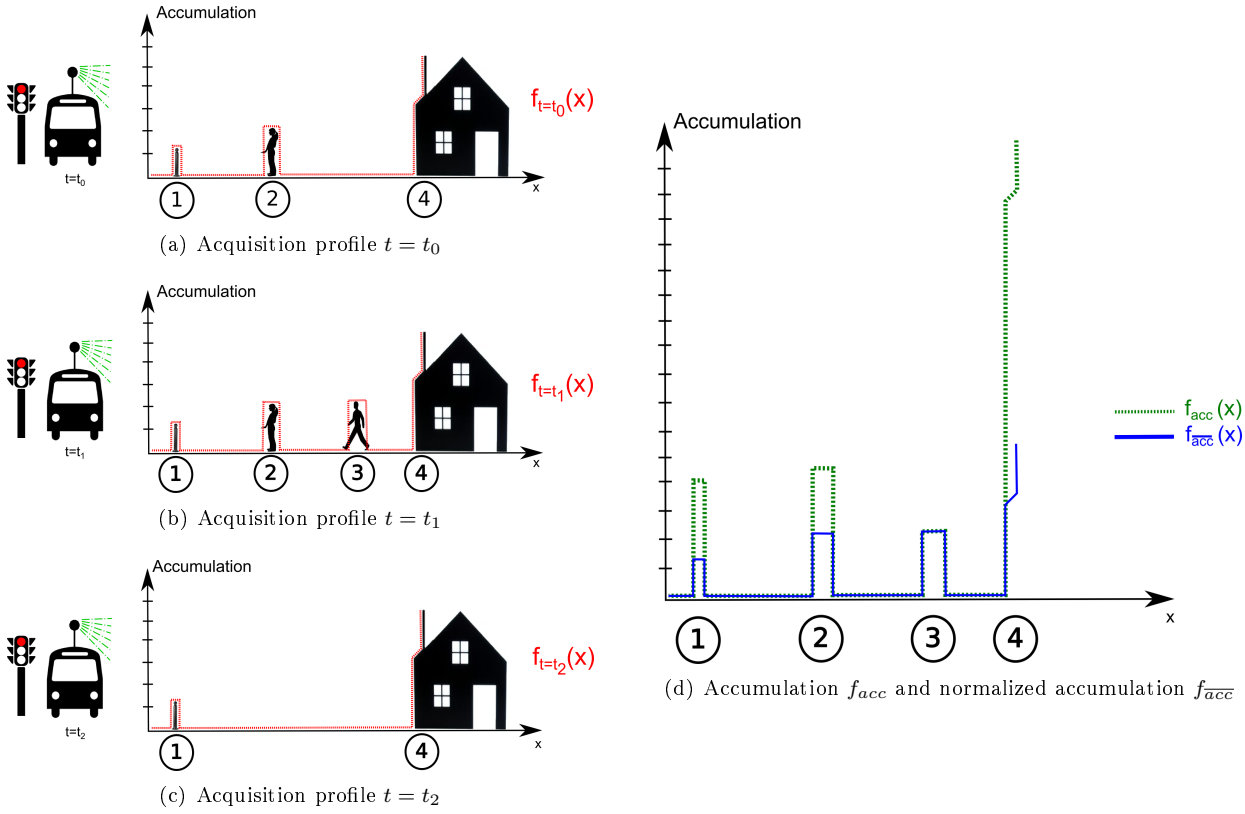
(c) Histogram of radius before filtering: radius ∈ (2.4, 453.9) m. Linear (blue) and Log (cyan) scale.

(d) Histogram of radius after filtering: radius ∈ (2.4, 50.0) m. Linear (blue) and Log (cyan) scale.

Figure 3.12: Distant points filtering. Test site: *rue d'Assas*, Paris. Acquired by Stereopolis II, IGN©France. In order to filter out distant points, a simple threshold is applied to the *radius* of each point. According to technical and practical issues, points farther than 50 m are not considered. Note that the elevation image computed from the filtered point cloud (a) contains approximately 30 times less pixels than that computed from the original point cloud (b). Analyzing their histograms (c) and (d), respectively, we can see that relevant information is kept in the filtered point cloud.

on crossing the laser beam, as shown in Figure 3.13(d). In order to filter out redundant information at each point, only the profile providing the maximum accumulation is taken into account, as shown in Figure 3.13(d). This filtering led us also to define a normalization of the accumulation image $f_{\overline{acc}}$, which contains more reliable information.

Figure 3.14 shows an example of this phenomenon in a test site in *St. Sulpice* square in Paris. It is noteworthy that several misleading great accumulation values appear when the acquisition vehicle takes the turn around the square or when it stops. Figure 3.14(b) shows the maximal elevation image, while Figure 3.14(c) presents the labelling of each scan line in order to detect overlapping profiles. Using this strategy, redundant information is filtered out and more reliable results are obtained, as shown in the accumulation image of Figure 3.14(e).



(a) Acquisition profile $t = t_0$

(b) Acquisition profile $t = t_1$

(c) Acquisition profile $t = t_2$

(d) Accumulation f_{acc} and normalized accumulation $f_{\overline{acc}}$

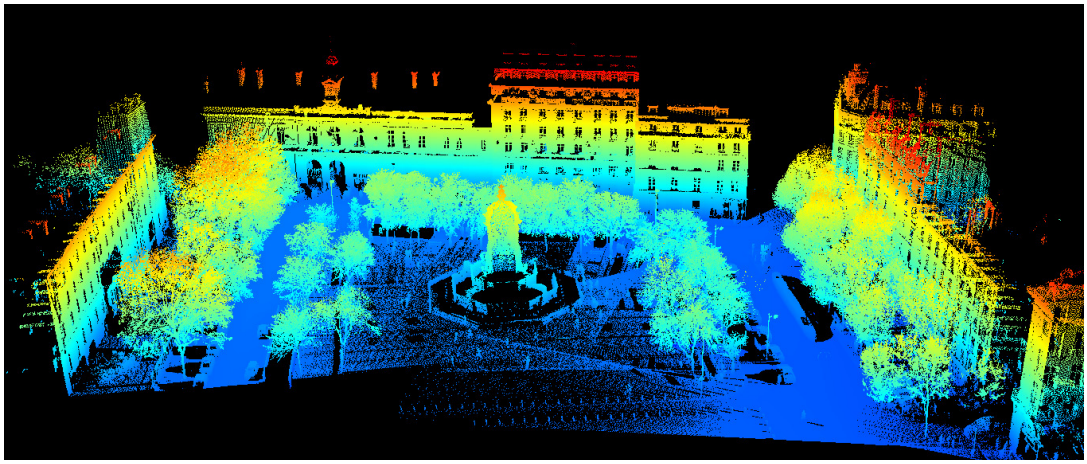
Figure 3.13: Filtering redundant information on overlapping profiles. Consider the acquisition vehicle stopped during the interval $[t_0, t_2]$ due to a traffic light and the three overlapping scan lines $f_{t=t_0}(x)$, $f_{t=t_1}(x)$ and $f_{t=t_2}(x)$, as shown in (a), (b) and (c). The urban scene contains the following urban objects enumerated from ① to ④: ① bollard, ② pedestrian, ③ pedestrian and ④ house facade. On the one hand, the two fix objects ① and ④ appear in the three profiles, then their accumulation f_{acc} is three times the accumulation of an individual profile, as shown in (d). On the other hand, the two mobile objects ② and ③ appear in some profiles, then their accumulation f_{acc} not only depends on the object geometry but also on the time spent by the pedestrians on crossing the laser beam, as shown in (d). In order to filter out redundant information at each point, only the profile providing the maximum accumulation is taken into account, as shown in (d). This filtering led us also to define a normalization of the accumulation image $f_{\overline{acc}}$, which contains more reliable information.

3.6.3 Image interpolation

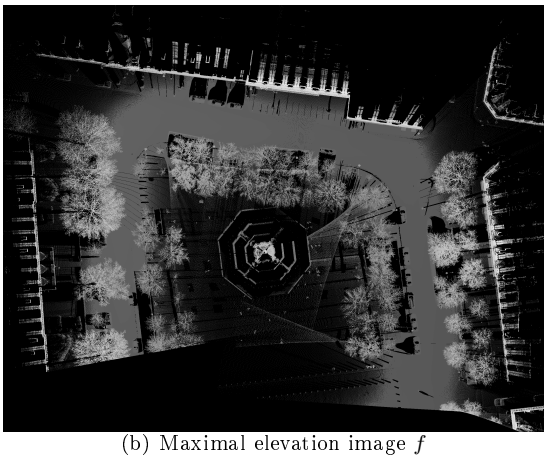
After projection to elevation images, an interpolation is required in order to fill holes caused by occlusions and missing scan lines. A morphological interpolation based on filling holes is chosen since this transformation does not create new regional maxima, it can fill holes of any size and no parameters are required. This is important in order to avoid false alarms in the object detection approach (explained later in Chapter 6).

In the most simple sense, a hole is a dark region (regional minimum) which is not connected to the image border and is surrounded by brighter pixels. The fill–holes transformation is implemented as the reconstruction by erosion ($R^{\epsilon}_{f}(f_{marker})$) of image f from marker f_{marker}, as shown in Figure 3.15. Marker f_{marker} is set to the maximum image value everywhere except along the image border, where the original image value is kept. Applying this transformation, each hole is filled with the lowest value in its boundary. Figure 3.15(b) illustrates this definition on the 1D case. Note that a minimum in the left part of the signal is not a hole because it touches the border. In order to preserve original data, only pixels with no data are modified, while other pixels keep their original value. For further details on gray-scale reconstruction operators, the reader should refer to (Soille and Ansoult, 1990; Vincent, 1993).

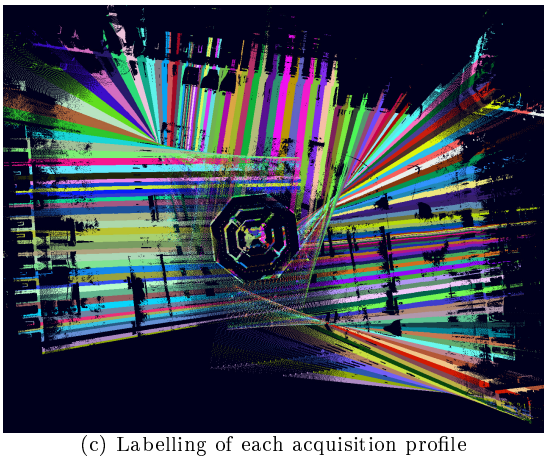
Let us explain our interpolation method with an example. Figure 3.16 illustrates a typical acquisition profile. The urban profile contains the following urban objects enumerated from ① to ⑦: ① car, ② pedestrian, ③



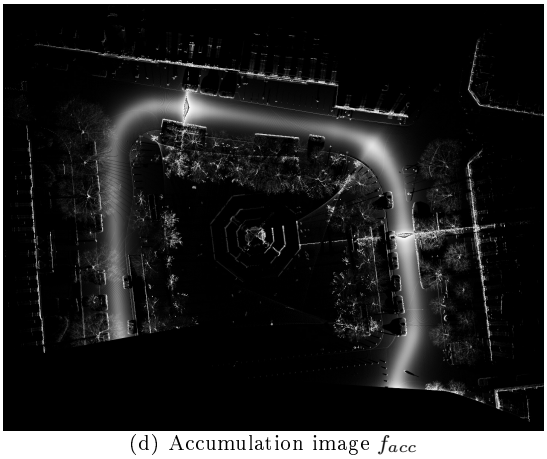
(a) 3D point cloud colored with the Z coordinate



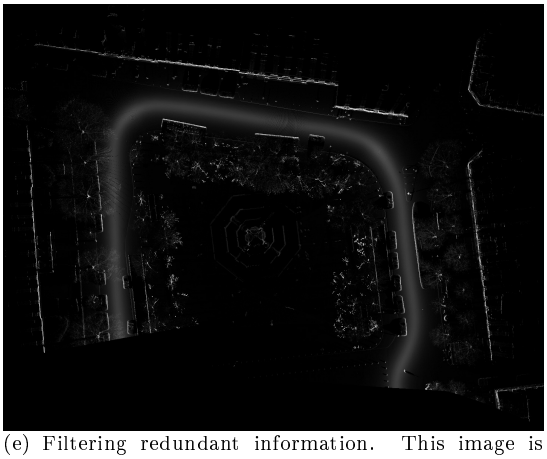
(b) Maximal elevation image f



(c) Labelling of each acquisition profile

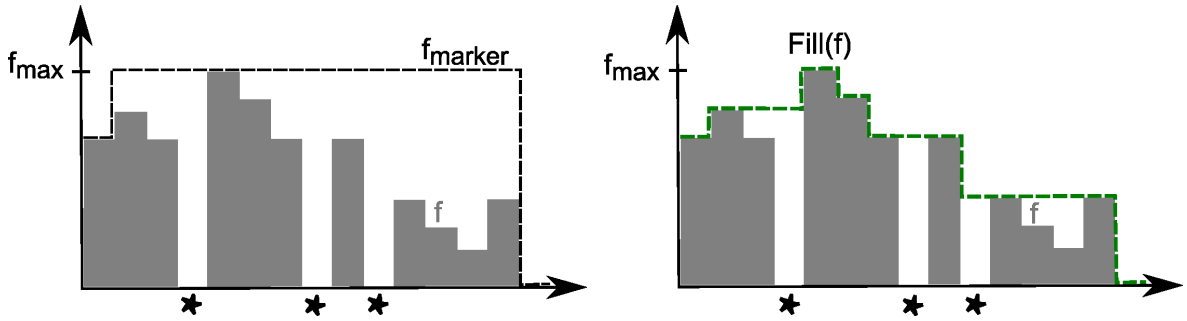


(d) Accumulation image f_{acc}



(e) Filtering redundant information. This image is called normalized accumulation image $f_{\overline{acc}}$

Figure 3.14: Filtering redundant information using the accumulation information. Test site in *St. Sulpice* square in Paris, France. Data acquired by Stereopolis II, IGN©France. It is noteworthy that several misleading great accumulation values appear when the acquisition vehicle takes the turn around the square or when it stops. (b) shows the maximal elevation image, while (c) presents the labeling of each scan line in order to detect overlapping profiles. Using our filtering strategy, redundant information is filtered out and more reliable results are obtained (e).



(a) Input function f and marker f_{marker}

(b) $Fill(f) = \hat{f} = R^{\epsilon}_{f}(f_{marker})$

Figure 3.15: Fill–holes transformation. In the most simple sense, a hole is a dark region (regional minimum) which is not connected to the image border and is surrounded by brighter pixels. The fill–holes transformation is implemented as the reconstruction by erosion ($R^{\epsilon}_{f}(f_{marker})$) of image f from marker f_{marker}. Marker f_{marker} is set to the maximum image value everywhere except along the image border, where the original image value is kept. Applying this transformation, each hole is filled with the lowest value in its boundary. Note that a minimum in the left part of the signal is not a hole because it touches the border. In order to preserve original data, only pixels with no data are modified ($\star$ indicates no–data points), while other pixels keep their original value.

noisy structure, ④ dog, ⑤ pedestrian, ⑥ house facade, and ⑦ chimney. Note that this is only an illustrative example on the 1D case. The processing is performed on the entire 2.5D elevation image. Using our interpolation method, each hole is filled with the minimal value surrounding the hole. For example, consider the hole in the left part, between objects ③ and ④. This hole is filled at the ground level because in 2.5D it is connected to ground pixels. Additionally, consider the holes in the left part, between objects ② and ③, and in the right part, between objects ⑤ and ⑥. These holes are also filled at the ground level even if the ground is not the minimal surrounding value in this 1D profile. We assume that these holes can be filled at that level because the ground is not occluded by the pedestrians ② and ⑤ in the previous or in the following profiles.

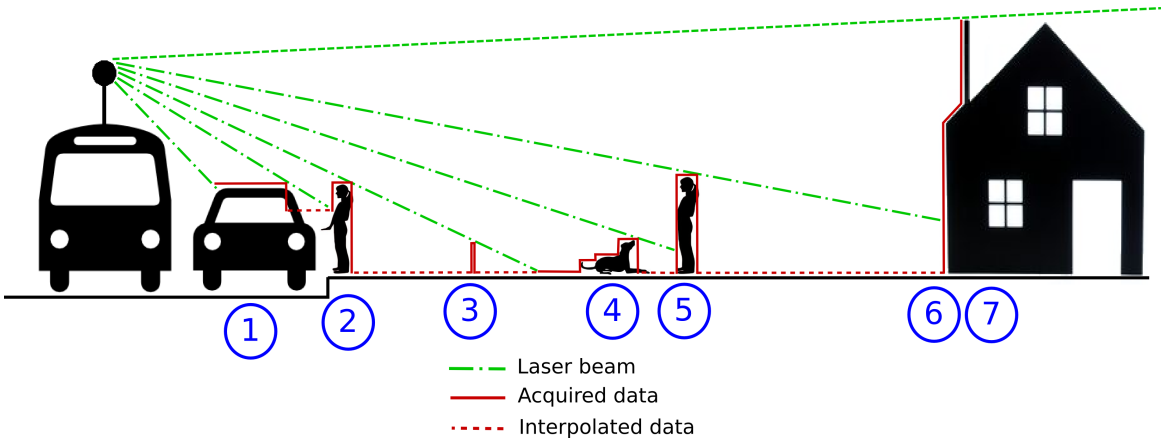


Figure 3.16: Acquisition scheme and interpolation method on the 1D case. The urban profile contains the following seven urban objects: ① car, ② pedestrian, ③ noisy structure, ④ dog, ⑤ pedestrian, ⑥ house facade, and ⑦ chimney.

Figure 3.17(a) exhibits an experimental scenario in *St. Sulpice* square in Paris. Figure 3.17(b) presents the elevation image f, where black pixels indicate no data. In the middle of the square, sparse points are obtained as the result of occlusions and faraway objects. This image has to be interpolated before processing. Note that almost all dark regions are touching the image border, so they would not be filled by a classical fill–holes transformation. To solve this problem, each isolated region is connected to its closest neighbor by the shortest

path. Next, the artificial boundaries for each influence zone are defined using the maximal value on the elevation image, as shown in Figure 3.17(c). Then, the fill-holes transformation is applied in order to interpolate the image (Figure 3.17(d)). Finally, the maximal elevation value on the artificial boundaries is replaced by the result of a morphological opening. Note that our methodology performs well on near objects. However, several false artifacts can appear when interpolating distant objects because there are not enough points. This can be easily corrected in the segmentation step eliminating objects for which the number of interpolated points ($N_{\text{interpolated}}$) is much higher than the number of acquired points (N_{acquired}). That leads us to define the confidence index C presented in Equation (3.2). This index is useful to characterize urban objects, as it will be shown later in the classification approach presented in Chapter 6. The distance from a point to the acquisition system could also be considered.

$$C = \frac{N_{\text{acquired}}}{N_{\text{acquired}} + N_{\text{interpolated}}} \tag{3.2}$$

This interpolation is fast, interpolates holes of any size and is parameterless. However, one of the disadvantages is that holes are filled by flat zones, *i.e.* all the pixels in the hole will have the same elevation after interpolation. Thus, it may not give realistic results in the case of steep terrains, as shown in Figure 3.10.

3.7 Conclusions

In order to process 3D urban data, we proposed the use of elevation images since they are convenient structures to visualize and to process data using all the large collection of image processing tools, specially mathematical morphology. Projecting 3D information to images implies a reduction in the amount of data to be processed with respect to the input 3D point cloud. Moreover, neighborhood relationships in the elevation image are intrinsically defined. In other words, we process an elevation image using image processing techniques, which is much faster than processing the 3D points directly.

During the projection, the number of pixels per length unit k is the only free parameter. It controls the elevation image size and it has to be carefully chosen. On the one hand, if k is too small, fine details are not preserved because too many points would be projected on the same pixel. On the other hand, too large k presuppose connectivity problems and large image sizes, which implies high computational time. This analysis leads us to prospect a multi–scale approach where initial structures are detected at low scales using a fast approach (small k), and then a refinement may be carried out in higher levels (large k) in order to get more accurate results.

Additionally, distant and isolated points in the 3D point cloud have a critical impact since they can unnecessarily increase the elevation image size. In order to filter out not reliable and distant points, a thresholding is applied to the *radius* of each point. If the distance from the laser to the 3D point is greater than 50 m, that point is not considered. In general, such distant points are located in road intersections and come from streets perpendicular to the vehicle trajectory. In a large scale application, it is supposed to have a complete survey of the urban scenario. Thus, perpendicular streets are supposed to be mapped later by the vehicle. Therefore, too distant points can be eliminated in a harmless way.

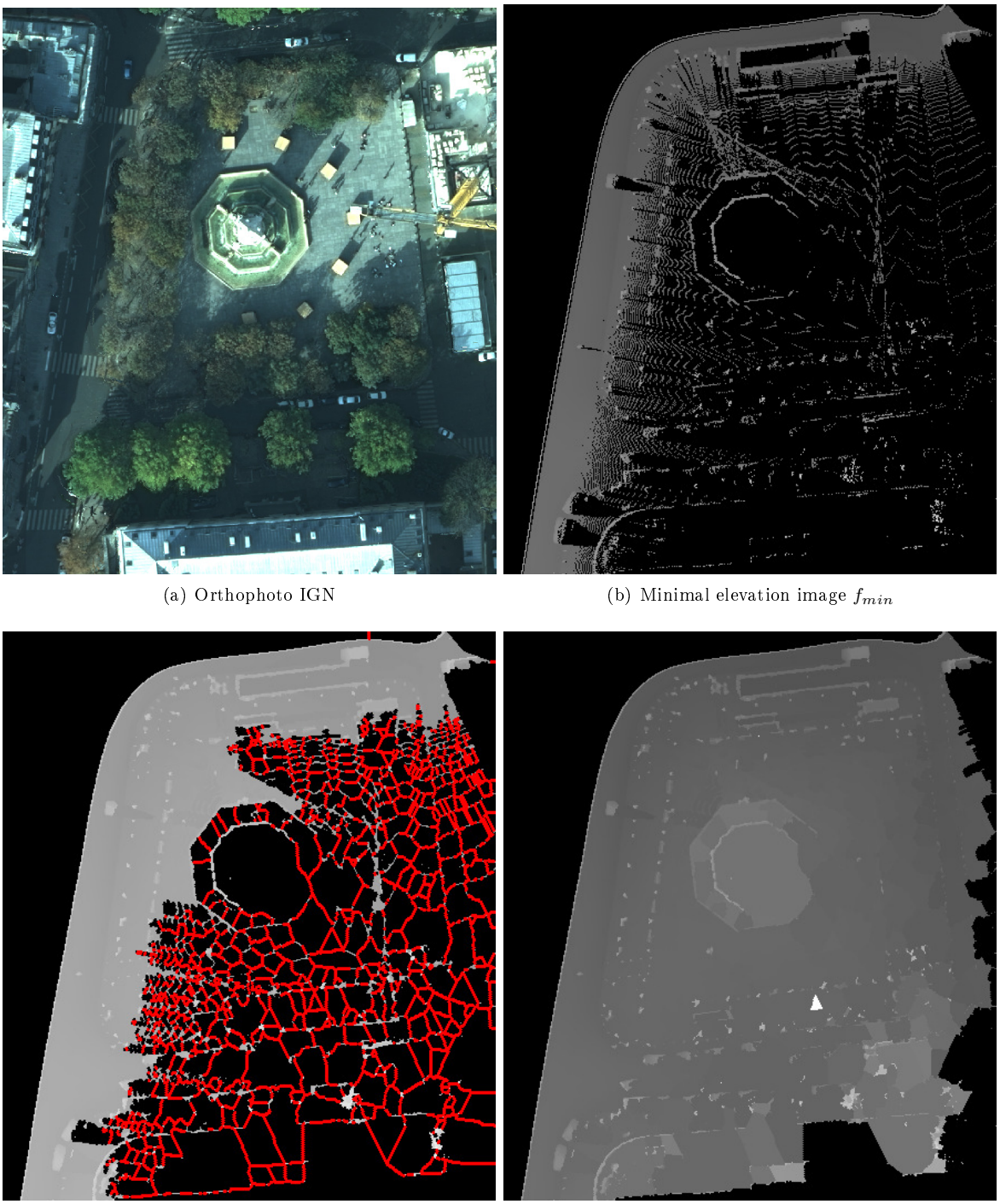
One of the disadvantages of processing 3D urban data using a single elevation image is that high objects such as trees may occlude objects below them. To solve this problem, we propose a projection strategy using slices. In our experiments, we have defined two slices for obstacles lower than and higher than 3.5 m, respectively. This processing can be interpreted as an adaptive voxelization where the dimensions of each voxel are determined by the image pixel size and the slice height, and the voxel location is determined by the ground.

In general, several points can be projected on the same pixel. Then, an accumulation image is defined storing the number of points projected on each pixel. This information is very useful to detect vertical high structures in the scene. Moreover, it is useful to identify redundant information since overlapping scan lines give high accumulation values. Redundant information should be properly managed in order to get reliable results and to reduce processing time. In order to reduce redundant information, we propose a filtering strategy using the accumulation. When two or more acquisition profiles overlap, only the profile providing the maximal quantity of information is kept. This filtering is carried out pixel by pixel on the accumulation image, thus the result is a composition of segments from individual profiles.

In order to manage occlusions and connectivity problems due to missing scan lines, a morphological interpolation based on filling holes is applied. This transformation is fast, does not create new regional maxima, can fill holes of any size and does not require parameters. One of the disadvantages is that holes are filled by a

flat zone, *i.e.* all the pixels in the hole will have the same elevation after interpolation. This strategy may give non-realistic results in the case of steep terrain. Using several scans of the same zone, velodyne sensors and panoramic color images can help to reduce occlusion problems.

Although the idea of deriving elevation images from 3D point clouds is not new, the development of accurate and fast preprocessing algorithms is still an open problem in the scientific community. In particular, the works by Hernández and Marcotegui (2009c) and Hernández (2009) have been the starting point of this thesis. Their projection and interpolation methods (Section 3.4.1 and Section 3.6.3) have been applied identically as in their works. Nevertheless, the projection by slices and the filtering methods proposed here are entirely original and constitute one of the main contributions of this chapter.



(a) Orthophoto IGN

(b) Minimal elevation image f_{min}

(c) Artificial boundaries (red)

(d) Interpolated image $\hat{f}_{min}$

Figure 3.17: Image interpolation. (a) exhibits an experimental scenario in *St. Sulpice* square in Paris. (b) presents the elevation image f, where black pixels indicate no data. Note that almost all dark regions are touching the image border, so they would not be filled by a classical fill–holes transformation. To solve this problem, each isolated region is connected to its closest neighbor by the shortest path. Next, the artificial boundaries for each influence zone are defined using the maximal value on the elevation image, as shown in (c). Then, the fill-holes transformation is applied in order to interpolate the image (d). Finally, the maximal elevation value on the artificial boundaries is replaced by the result of a morphological opening.

4 Ground segmentation and accessibility analysis

4.1 Résumé

Dans ce chapitre, nous présenterons deux contributions de cette thèse sur la segmentation automatique du sol et l'analyse d'accessibilité urbaine à partir de données 3D. Dans un premier temps, nous présenterons une révision de l'état de l'art. Dans un deuxième temps, nous exposerons la segmentation du sol basée sur des images d'élévation ainsi que notre extraction et caractérisation des bords de trottoirs à partir de critères géométriques et contextuels. Finalement, nous reporterons des résultats quantitatifs sur des bases de données disponibles dans la littérature.

4.2 Introduction

In general, large cities are built according to demographical and geographical constraints, architectural preferences and governmental budgets. Many places are not accessible for wheelchairs, skaters, segways, baby buggies, among others. Taking a short trip through the street can reveal an environment plagued with physical barriers, obstacles, narrow sidewalks and inappropriate ramp access, where wheelchair users bear the brunt.

Contrarily to the general idea, accessibility affects not only disabled persons but also old people, children and pregnant women, as shown in Figure 4.1. It is noteworthy that 46% of people is concerned by accessibility in urban areas.

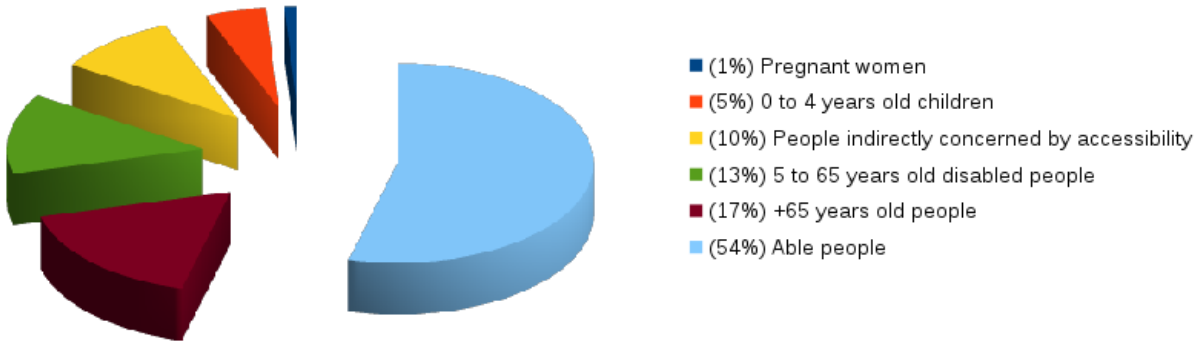


Figure 4.1: People concerned by accessibility in urban environments. Contrarily to the general idea, accessibility affects not only disabled persons but also old people, children and pregnant women. It is noteworthy that 46% of people is concerned by accessibility in urban areas.

About 80 million people living in European Union (EU) have a mild to severe disability. Physical obstacles make them vulnerable to social exclusion, low employment and limited education level. In fact, mean poverty rate for those with disabilities is 70% higher than the average. In 2007, EU signed the *United Nations convention on the rights of persons with disabilities* (UN, 2007). The aim is allowing people with disabilities to go on their daily lives like everyone else and enjoy their rights as EU citizens. One of the strategies consists in ensuring physical access to buildings, roads, transportation, schools, housing, medical centers and workplaces.

In France according to Law 2005-102[1], local authorities are required to guarantee accessibility to public spaces. Thus, it is very important to be able to make large scale accessibility diagnoses in urban environments in order to identify places requiring adaptation. An available manual–assisted solution is Wheelmap (Sozialhelden, 2012).

[1] Loi 2005-102 du 11 février 2005: "Pour l'égalité des droits et des chances, la participation et la citoyenneté des personnes handicapées".

It is an on-line service, based on Open Street Maps, aiming at tagging wheelchair–accessible places. It is a crowd–sourcing project where everyone can collaborate by tagging public places according to accessibility for persons using wheelchairs. Other solutions include automatic 3D urban analysis techniques (Golovinskiy et al., 2009; Hernández and Marcotegui, 2009c; Pu et al., 2011; Douillard et al., 2011; Rutzinger et al., 2011). However, automatic urban accessibility analysis is still an open problem.

Urban accessibility information can be integrated into navigation services as on–line maps (Sozialhelden, 2012), support systems using cell phones (Rashid et al., 2010), collaborative social networks (Menkens et al., 2011), or even in automated wheelchairs and segways (García et al., 2010). Including accessibility parameters in city maps allows to define adaptive itineraries according to detailed and accurate information about barriers and obstacles in the public infrastructure.

As aforementioned in Chapter 1, our work is part of Cap Digital Business Cluster TerraMobilita project: "3D mapping of roads and urban public space, accessibility and soft-mobility". This project responds precisely to requests about 3D urban maps and soft–mobility applications. There are two general aims in this project: i) to develop new methods and tools to create and update urban maps using laser scanning and digital imagery. ii) to develop innovative applications for soft–mobility itinerary planning.

The contribution of the present chapter is twofold: automatic ground segmentation and curb accessibility analysis.

In the first part, our method segment ground, facades and objects in order to build 3D obstacle maps useful for itinerary planning. In particular, the work by Hernández and Marcotegui (2009a) has been the starting point of this chapter. Their segmentation method has been adapted in order to segment pavement (Section 4.4).

The second part is entirely original and constitute one of the main contributions of this chapter. It consists in developing automatic methods for segmentation, reconnection and characterization of curbs (Section 4.5) and urban accessibility analysis (Section 4.7). This constitutes one of the most attractive contributions of this thesis due to its social impact since urban accessibility affects not only disabled persons but also old people, children and pregnant women. In the framework of the *United Nations convention on the rights of persons with disabilities*, local authorities are required to guarantee accessibility in public spaces in order to reduce social exclusion, low employment and limited education of people concerned by accessibility. One of our publications on this topic (Serna and Marcotegui, 2013b) has been awarded with the U. V. Helava Award for the 2013 best paper in the *International Society for Photogrammetry and Remote Sensing* (ISPRS Journal volumes 75-86) http://www.isprs.org/society/awards/helava/2013.aspx.

For our experiments, 3D laser scanning data are acquired by Stereopolis II (Paparoditis et al., 2012) and L3D2 (Goulette et al., 2006a), two mobile laser scanning (MLS) systems from IGN France and MINES ParisTech, respectively. Additionally, a public database from Enschede (The Netherlands) is used to get quantitative results and to compare our methods with the state of the art. Technical details on TerraMobilita acquisition systems and Enschede database can be found in Section 2.4 and Section 2.5.4, respectively.

This chapter is organized as follows. Section 4.3 reviews related works in the state of the art and establishes their differences with respect to our work. Section 4.4 describes our ground segmentation method based on elevation images. Section 4.5 introduces our methods for curb segmentation and reconnection. Section 4.7 explains the accessibility analysis and presents an illustrative itinerary planning application. Section 4.8 presents quantitative results with respect to other databases available in the literature. Finally, Section 4.9 concludes this chapter.

4.3 Related work

Bab-Hadiashar and Gheissari (2006) propose a method to segment planar and curved surfaces in range images. Their method consists in selecting the appropriate parametric model that minimizes strain energy of fitted surfaces. The authors applied their methodology to indoor range images of the University of South Florida (USF) database (Hoover, 1994). This work can be extended in order to segment surfaces such as ground and facades on elevation images. Several works on the parametric model fitting problem can be found in the literature (Boyer et al., 1994; Werghi et al., 1998; Marshall et al., 2001; Chaperon and Goulette, 2001). The main drawback of these methods is that they involve the model selection problem which can be different for different images, are time consuming due to minimization procedures and may produce sub-segmentation.

Ayres and Kelkar (2006) present a method to characterize sidewalks based on ground elevation profiles. The aim is identifying unusual elevation changes and obstacles. The authors highlight the impact of elevation changes for pedestrians and soft–mobility users safety. They work on very precise data from streets in California. Unfortunately, this is not suitable for large scale applications since data are acquired manually.

Hernández and Marcotegui (2009a) use elevation images from 3D point clouds in order to extract quasi–flat zones on the ground and use them as markers for a constrained watershed (Beucher and Meyer, 1993). Then, a region adjacency graph is used to determine the border between roads and sidewalks. This procedure fails in presence of access ramps, as shown in Figure 4.9(a). In that case, quasi-flat zones merge roads and sidewalks and it is no longer possible to detect curbs.

Vosselman and Zhou (2009) use Aerial Laser Scanning (ALS) data to detect curbstones. First, small height jumps at ground level are detected. Second, a smooth curve is fitted to generate the separation between sidewalk and road. Finally, small gaps between nearby and collinear segments are closed. In unoccluded regions, this approach is robust to traffic signs, cars and multi–line roads. Good results are presented with respect to ground truth measurements. However, airborne laser resolution (20 points/m^2) is not precise enough to detect access ramps and low curbs. In fact, only curbstones of at least 10 cm height can be clearly detected. Denis et al. (2010) adapt the work by Vosselman and Zhou (2009) to MLS data in order to extract and model urban roads as 3D surfaces. In that work, authors combine MLS data and road axes derived from aerial imagery. Ground points are extracted with an adapted surface growing method. Then, curbs are detected based on the elevation gradient on a strip-by-strip basis. Recently, Zhou and Vosselman (2012) extended their methodology to denser (1000 points/m^2) MLS datasets. They improve the planimetric accuracy of curbstone locations fitting a sigmoidal function to detected points, in a similar way to Siegemund et al. (2010). In such methods, although detection is performed directly on the 3D point cloud, it is made on a strip by strip basis, so intrinsic information between neighboring strips is missing. More recently, Serna and Marcotegui (2013b) solved this problem by processing all strips at the same time using elevation images.

Valero et al. (2010) present an automatic road extraction methodology for high resolution imagery. This procedure can be extrapolated to curb detection since both problems have similar assumptions: curbs are thin and elongated paths, but not necessary straight, and present color differences with respect to their neighborhood. Their experimental results show accurate road extractions in terms of completeness and correctness. Some post processing has to be proposed in order to reconnect isolated segments due to occlusions and shadows.

Gang and Guangshun (2010) propose an approach to model urban road networks based on manual markers. They use an interactive interface to mark sidewalks and roads on aerial images. Then, Bézier curves and polygons are used to model the road. This method is realistic and very fast to render, however manual marking is a time consuming task. Automatic detections and road network databases are needed for large scale modeling.

Hervieu and Soheilian (2013a,b) investigate the surface modeling of roadways and pavements from MLS data. First, road border detection is considered. A system recognizing curbs and access ramps while reconstructing the missing information in case of occlusion is presented. A user interface scheme is also developed, providing an effective tool for semi-automatic processing of large amount of data. Then, based upon road edge information, road and pavement surfaces are reconstructed. The main drawback of this method is that manual intervention may be time-consuming. Automatic methods segmenting curbs can be introduced at the process input in order to improve both performance and speed.

With respect to other works reviewed in the state of the art, we aim at developing an automatic method for ground segmentation and curb accessibility diagnosis suitable for large scale applications. Our method is automatic using few constraints on 3D laser scanning data, creates an obstacle map segmenting objects and facades, defines the accessibility for each curb, and can manage reconnection problems due to access ramps and occlusions. A detailed description is presented in the following sections.

4.4 Ground segmentation

Ground segmentation is one of the most important steps in urban semantic analysis since it allows the creation of digital terrain models (DTM). In the case of accessibility diagnosis, ground segmentation is a critical step since curbs and obstacles are located on it.

Figure 4.2 shows our proposed methodology. First, input point cloud is mapped to elevation images and a morphological interpolation is applied, as explained before in Chapter 3. Second, the quasi–flat zones algorithm is used to segment the ground, including roads and sidewalks. Third, facades and objects are segmented using morphological transformations (details will be presented in Chapters 5 and 6) and the obstacle map is defined. Fourth, curb candidates are segmented using height and elongation criteria, and close curbs are reconnected using Bézier curves based on semantic information. Then, curb accessibility is defined according to international standards. Finally, the obstacle map and the accessibility information can be exported to a Geographical Information System (GIS) and used to define adaptive itineraries according to different types of soft-mobility.

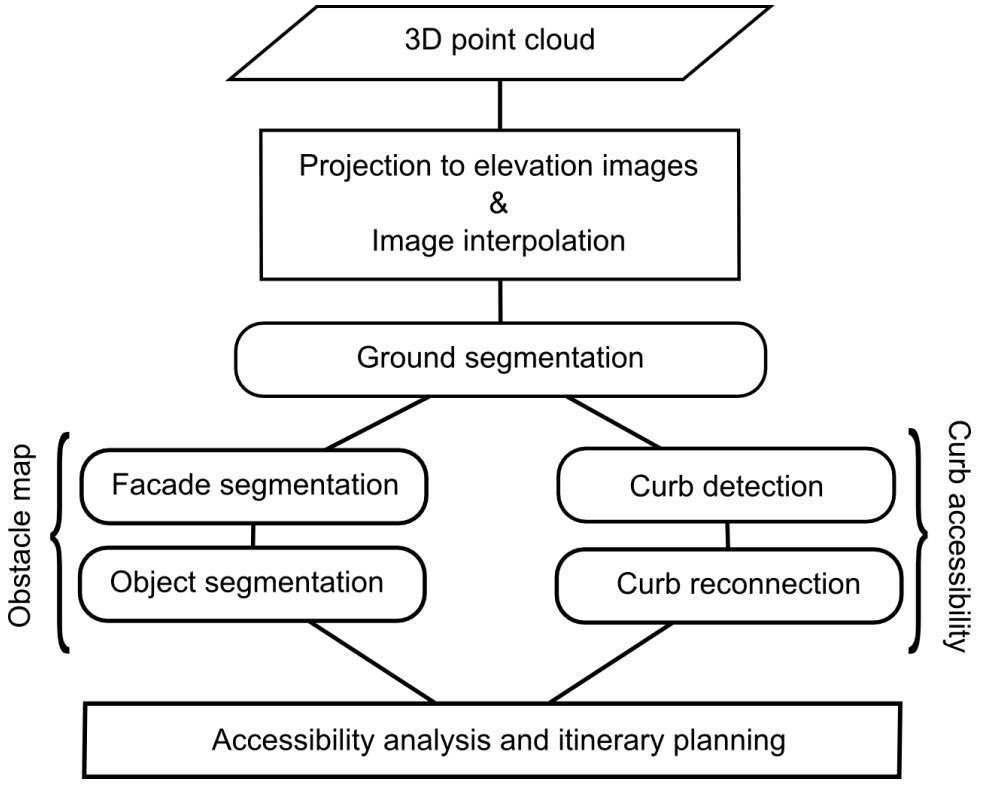


Figure 4.2: Work–flow of our proposed urban accessibility analysis from 3D laser scanning data. First, input point cloud is mapped to elevation images and a morphological interpolation is applied, as explained before in Chapter 3. Second, the quasi–flat zones algorithm is used to segment the ground, including roads and sidewalks. Third, facades and objects are segmented using morphological transformations (details will be presented in Chapters 5 and 6) and the obstacle map is defined. Fourth, curb candidates are segmented using height and elongation criteria, and close curbs are reconnected using Bézier curves based on semantic information. Then, curb accessibility is defined according to international standards. Finally, the obstacle map and the accessibility information can be exported to a Geographical Information System (GIS) and used to define adaptive itineraries according to different types of soft-mobility.

4.4.1 The quasi-flat zones algorithm

Several methods found in the literature try to solve the ground segmentation problem fitting horizontal planes to the 3D point cloud (Goulette et al., 2006a; Gorte, 2007; Schnabel et al., 2008; Zhu et al., 2010; Pu et al., 2011; Poreba and Goulette, 2012b; Aijazi et al., 2013). The main drawback of these approaches is that they may fail in urban scenarios due to non-flat roads, access ramps and speed humps.

To solve this problem, we propose a segmentation method taking advantage of the quasi-flat character of the ground. In particular, the work by Hernández and Marcotegui (2009a) has been our starting point. Their segmentation method has been adapted in order to segment the ground. In the context of that work, their procedure may fail because the quasi-flat, called also λ-flat, zones algorithm is used to define a mask containing ground and objects while filtering facades. Afterwards, objects are filtered out and a ground mask is generated. In our method, we firstly filter out facades and objects. Thus, our ground segmentation method does not depend on the object detection result, providing more robust results. Moreover, we take advantage of the presence of access ramps to extract the complete ground mask merging roads and sidewalks. The method is as follows.

In order to segment the ground, the minimal elevation image f_{min} is used because it contains the lowest projected point on each pixel, which is generally the ground. This image is interpolated using a morphological based method, presented in Chapter 3, in order to avoid connectivity problems. Then, the λ-flat zones labeling algorithm is used. This algorithm was firstly introduced in image processing by Nagao et al. (1979) and was defined by Meyer (1998) as:

Definition 4.4.1 *Let f be a digital gray–scale image $f : D \rightarrow V$, with $D \subset Z^2$ the image domain and*

$V = [0, ..., R]$ the set of gray levels. Two neighboring pixels p, q belong to the same λ–flat zone of f, if their difference $|f_p - f_q|$ is smaller than or equal to a given λ value.

The definition of λ–flat zones is very useful in image partition, simplification and segmentation. However, it suffers from the well–known chaining effect of the single linkage clustering (Duda et al., 2000). That is, if two distinct image objects are separated by one or more transitions going in steps having a gray–level difference lower than λ, they will be merged in the same λ–flat zone.

In our urban analysis based on elevation images, this approach allows to segment the ground in despite of its curvature and slope. Using Definition 4.4.1, ground $\hat{f}_{gr}$ is obtained as the largest quasi-flat zone on the interpolated minimal elevation image $\hat{f}_{min}$.

Figure 4.3 shows an example of our ground segmentation method in a test site in *St. Sulpice square* in Paris, France. Figure 4.3(a) shows the minimal elevation image f_{min} and Figure 4.3(b) the interpolated minimal elevation image $\hat{f}_{min}$. It is noteworthy that ground is not perfectly flat, as shown in Figure 4.3(c), where each different color represents a different flat-zone, *i.e.* a maximal connected component of constant gray–level (Salembier and Serra, 1995).

Applying the quasi-flat zones labeling algorithm, ground is extracted as the largest λ–flat zone on $\hat{f}_{min}$ for a given λ parameter. Note that for $\lambda = 1$ cm (Figure 4.3(d)) several parts of the ground are missing while for $\lambda = 50$ cm (Figure 4.3(f)) the propagation is too permissive and several objects are reached by the propagation. In our experiments, we set $\lambda = 20$ cm because it is usually high enough to merge roads and sidewalks (even if there are no access ramps) without merging objects.

Figure 4.4 presents a result of our ground segmentation method on a 3D point cloud from *rue Cassette* in Paris, France. Note that this approach correctly segments the ground in spite of its curvature and the presence of a speed hump.

4.4.2 Obstacle map generation

Once the ground is extracted, all remaining structures are considered as facades and objects. Discrimination between them is important because facades define the public space boundary while urban objects define the obstacle map required for itinerary planning. On the one hand, facades are the highest vertical objects on the urban scene and they appear as elongated structures on interpolated maximal elevation image $\hat{f}$. Thus, they are segmented using morphological methods based on geometric and geodesic attributes. Facade segmentation is out of the scope of this chapter and it will be explained later in Chapter 5. Besides, several other works aiming at segmenting facades are available in the literature (Boulaassal et al., 2007; Hammoudi, 2011; Rutzinger et al., 2011; Poreba and Goulette, 2012b; Serna and Marcotegui, 2013a). On the other hand, urban objects appear as bumps and discontinuities on the ground on interpolated maximal elevation image $\hat{f}$. Thus, they are segmented using morphological methods based on the top–hat transformation by filling holes. Object segmentation is out of the scope of this chapter and it will be explained later in Chapter 6.

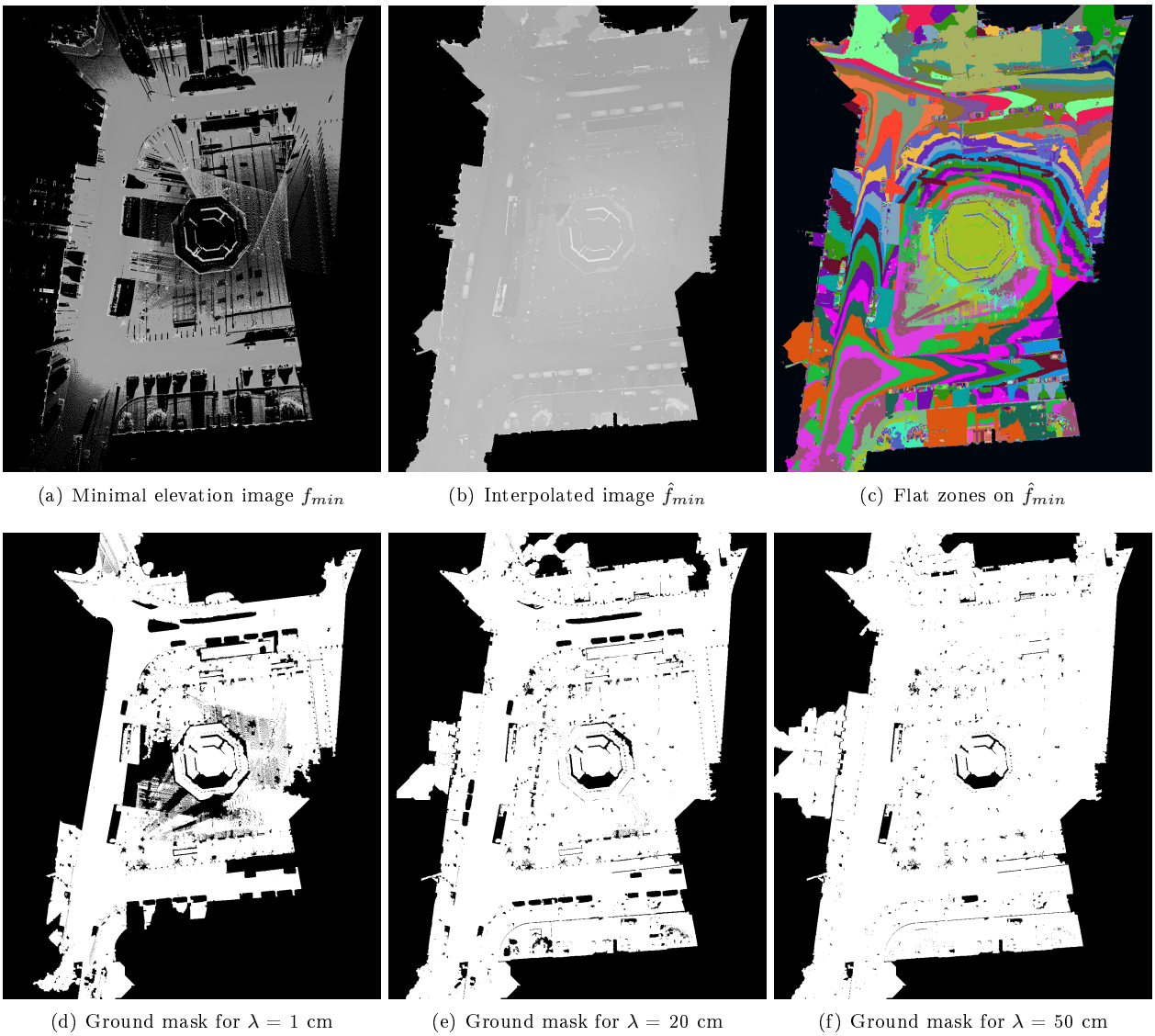
Figure 4.5 presents the 3D obstacle map obtained from ground, facade and object segmentation results. Note that all objects are assumed static. However, classification techniques can be used in order to distinguish mobile objects (*e.g.* pedestrians) from static ones (*e.g.* parked cars). For further information on object segmentation and classification methods, the reader is encouraged to review the Chapter 6 of the present thesis.

4.5 Curb segmentation and reconnection

Curb segmentation is very important in urban analysis applications since it defines the edge between roads and sidewalks. Besides, curb geometry is used to define the accessibility for a given type of mobility. Using the ground segmentation result, we propose a curb segmentation method based on the following geometrical hypothesis: "curbs are elevation discontinuities defining the limit between roads and sidewalks, and they appear as elongated edges on ground image $\hat{f}_{gr}$". Our proposed method is as follows:

First, the morphological external gradient is computed as the arithmetic difference between dilated ground image $\delta_B(\hat{f}_{gr})$ using a structuring element B and ground image $\hat{f}_{gr}$, as shown in Equation (4.1). In our experiments, a square structuring element of size 1 pixel is used. According to the spatial pixel size, which is an input parameter of our method discussed in Chapter 3, it corresponds to a dilation size between 5 and 10 cm. In order to avoid false alarms, gradients touching an interpolated zone are not considered.

$$\rho_{sup}(\hat{f}_{gr}) = \delta_B(\hat{f}_{gr}) - \hat{f}_{gr} \tag{4.1}$$



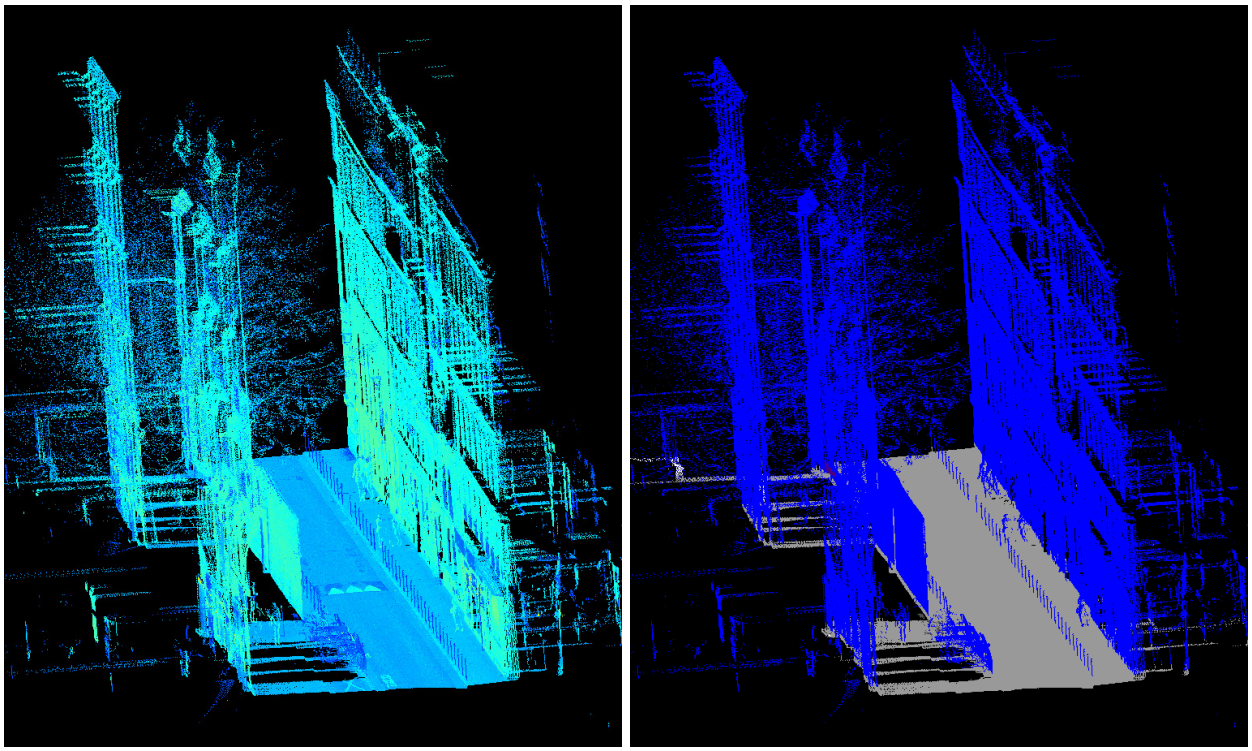
(a) Minimal elevation image f_{min} (b) Interpolated image $\hat{f}_{min}$ (c) Flat zones on $\hat{f}_{min}$

(d) Ground mask for $\lambda = 1$ cm (e) Ground mask for $\lambda = 20$ cm (f) Ground mask for $\lambda = 50$ cm

Figure 4.3: Ground segmentation: extraction of the largest quasi-flat (called also λ-flat) zone on interpolated minimal elevation image $\hat{f}_{min}$. The λ parameter chosen for our experiments is 20 cm. Test site in *St. Sulpice* square in Paris, France. Stereopolis II, IGN©.

Structures with elevation gradient between 3 and 20 cm are considered as curb candidates. Then, an elongation thinning is applied in order to filter out noisy and non-elongated structures. Experimentally, a minimum elongation E_{min}=10 has been defined in order to accept curbs. This threshold corresponds to the geodesic elongation of a curb of approximatively 1 m long and 0.08 m wide. We prefer geodesic measurements because curbs are usually not straight, so the Euclidean distance may sub-estimate their real length. For formal definitions and further details on the geodesic elongation, the reader is encouraged to read the Section 7.3.3 of this thesis.

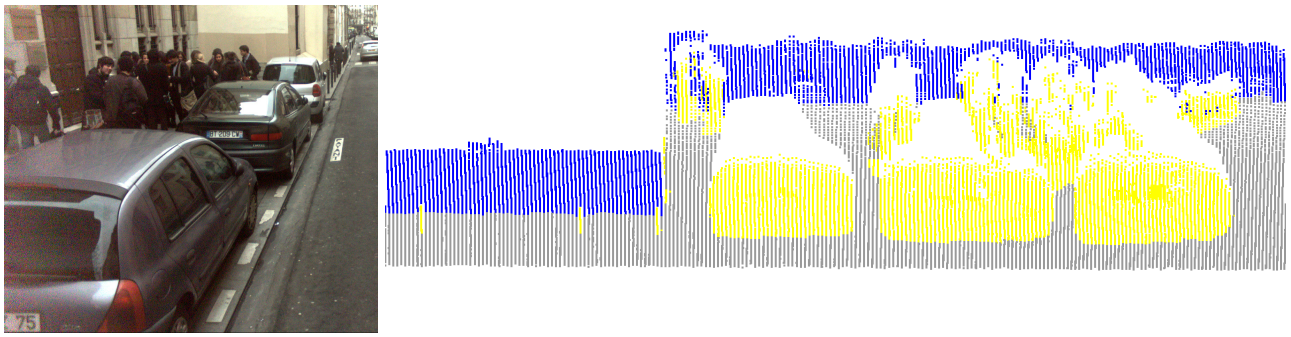
Figure 4.6 illustrates the effect of varying threshold E_{min} in our curb segmentation method. Note that E_{min}=0 (Figure 4.6(b)) preserves all structures between 3 and 20 cm height, E_{min}=5 (Figure 4.6(c)) does not take noise away and E_{min}=20 (Figure 4.6(d)) removes some real curbs, those that are short due to occlusions. Note that steps at building entrances are considered as curbs because their geometry hold our segmentation hypothesis. Their detection can be used to define building accessibility. In the case that they should not be considered, a constraint of minimal distance d_{facade} from the facade can be imposed.



(a) 3D point cloud colored by the reflectance.

(b) Segmentation result: ground (gray), other (blue).

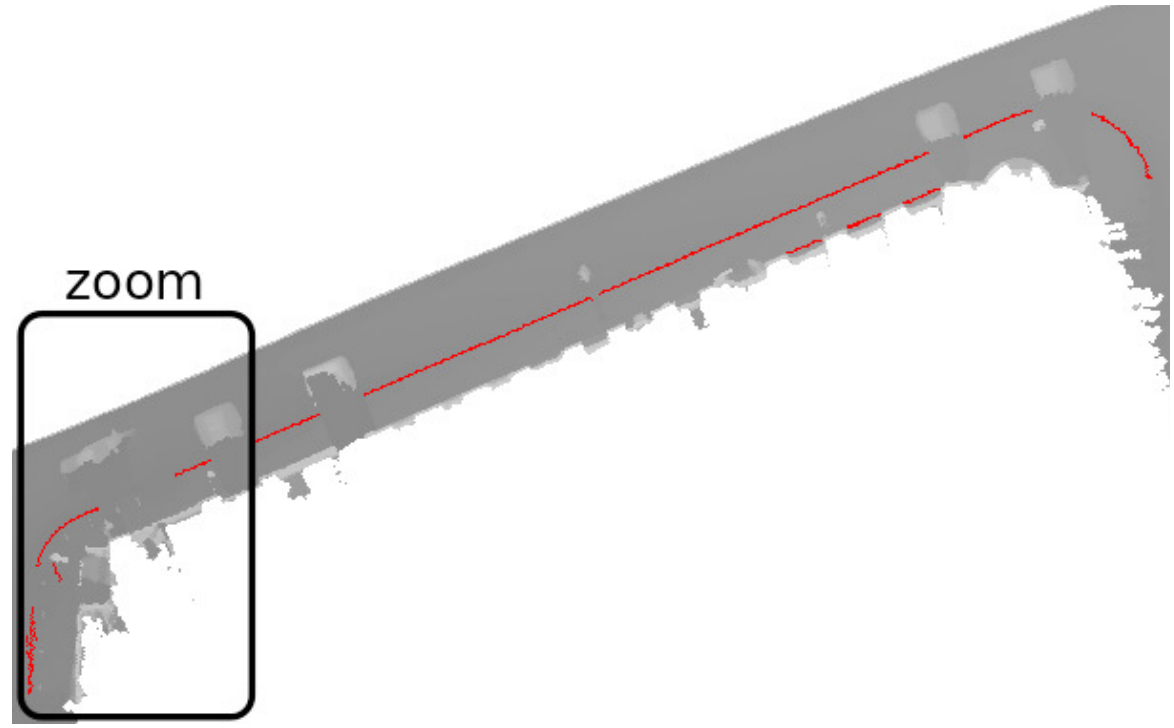
Figure 4.4: Ground segmentation result on a test site in *rue Cassette* in Paris, France. Stereopolis II, IGN©. Note that this approach correctly segments the ground in spite of its curvature and the presence of a speed hump.



(a) Illustrative photo

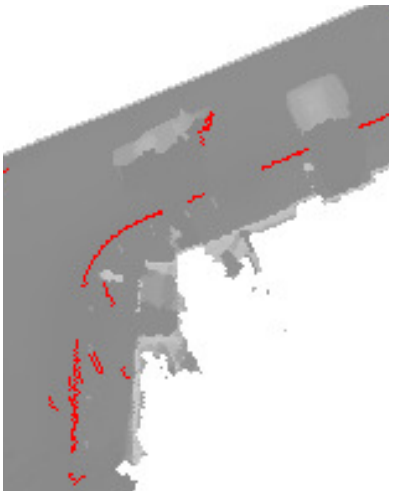
(b) 3D obstacle map: ground (gray), facades (blue) and objects (yellow).

Figure 4.5: Obstacle map generation in a test site in *rue Cassette* in Paris, France. Stereopolis II, IGN©. The 3D obstacle map is obtained from ground, facade and object segmentation results. Note that all objects are assumed static. However, classification techniques can be used in order to distinguish mobile objects (*e.g.* pedestrians) from static ones (*e.g.* parked cars). For further information on object segmentation and classification methods, the reader is encouraged to review the Chapter 6 of the present thesis.



(a) Curb detection for a geodesic thinning at E_{min}=10

(b) E_{min}=0



(c) E_{min}=5

(d) E_{min}=20

Figure 4.6: Curb segmentation for different geodesic elongation thinnings. In our experiments, E_{min}=10 has been chosen. Data acquired by IGN©France. Note that E_{min}=0 (b) preserves all structures between 3 and 20 cm height, E_{min}=5 (c) does not take noise away and E_{min}=20 (d) removes some real curbs, those that are short due to occlusions. Note that steps at building entrances are considered as curbs because their geometry hold our segmentation hypothesis. Their detection can be used to define building accessibility. In the case that they should not be considered, a constraint of minimal distance d_{facade} from the facade can be imposed.

4.5.1 Curb reconnection

The main drawback of this segmentation process is the lack of connectivity between curbs due to access ramps, occlusions, missing scan lines and acquisition problems. Some solutions can be found in the literature: Zhou and Vosselman (2012) close gaps between adjacent and co-linear curbs using lines; Shih and Cheng (2004) present an

approach based on adaptive mathematical morphology to link broken edges; and Talbot and Appleton (2007) propose a more sophisticated solution incorporating incomplete path openings. Unfortunately, these solutions are not suitable since reconnections through access ramps are not always straight (Figure 4.9(a)) and height discontinuities of access ramps are close to the noise level (Figure 4.9(c)).

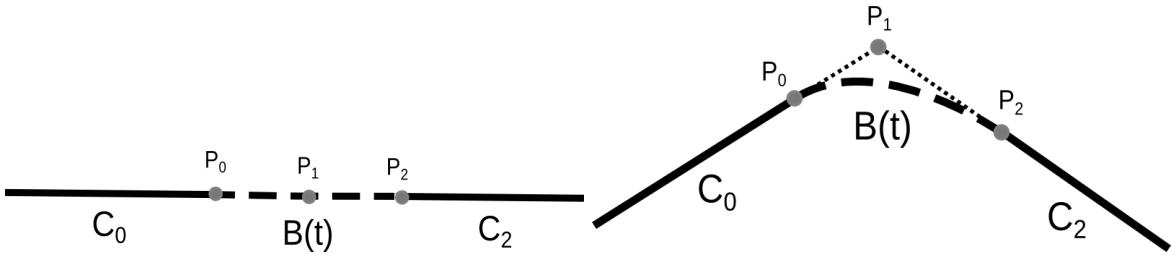
We propose a reconnection strategy based on quadratic Bézier curves. Two curbs closer than a given distance d_{min} are reconnected tracing a Bézier curve between their geodesic extremities (Morard et al., 2011a). Segment orientations are used in order to define Bézier curve parameters, as explained below.

A Bézier curve is a parametric path traced by function $B(t)$, given points P_0, P_1, and P_2, as shown in Equation (4.2). It departs from P_0 towards P_1, then bends to arrive to P_2. As a consequence, tangent lines in P_0 and P_2 both pass through P_1. Thus, the user can control input and output angles of the curve. This is an important smooth constraint, because in our application, initial and final angles of the reconnection should not change abruptly.

$$B(t) = (1-t)^2 P_0 + 2(1-t)tP_1 + t^2 P_2, \ \forall t \in [0,1] \tag{4.2}$$

Conveniently, the reconnection process can be written as the problem to find the three control points for a Bézier curve. Points P_0 and P_2 correspond to the geodesic extrema of curbs C_0 and C_2 to be reconnected. Thus, the problem is reduced to find the control point P_1. There are two types of reconnection using quadratic Bézier curves:

- If the curbs to be reconnected are co–linear, P_1 is put in the middle of the segment $\overline{P_0P_2}$. Therefore, the three control points are co–linear and the resulting reconnection is a straight line, as shown in Figure 4.7(a).
- If the curbs to be reconnected are not co–linear, P_1 is put in the intersection of the two projection lines from C_0 and C_2. Therefore, the resulting reconnection is a parabolic segment, as shown in Figure 4.7(b).



(a) Straight reconnection (b) Bent reconnection

Figure 4.7: Quadratic Bézier reconnection in straight and bent cases. (a) if the curbs to be reconnected are co–linear, P_1 is put in the middle of the segment $\overline{P_0P_2}$. Therefore, the resulting reconnection is a straight line. (b) if the curbs to be reconnected are not co–linear, P_1 is put in the intersection of the two projection lines from C_0 and C_2. Therefore, the resulting reconnection is a parabolic segment.

Let us explain our reconnection methodology with the toy example of Figure 4.8. First, Figure 4.8(a) shows three curbs A, B and C to be reconnected. Using a double propagation method from the barycenter, proposed by Morard et al. (2011a), geodesic extremities a_1, a_2, b_1, b_2, c_1 and c_2 are found. Second, for each geodesic extremity, the closest extremity belonging to a different curb is found. Since only reciprocal matches are allowed, only the matches (a_2, b_1) and (c_1, b_2) are candidates for reconnections. Third, if the Euclidean distance between matched extremities is shorter than a given threshold d_{min}, the reconnection becomes possible. In this toy example, we assume that Euclidean distances $d(a_2, b_1)$ and $d(c_1, b_2)$ are both shorter than or equal to d_{min}. In our experiments, this parameter has been adapted with respect to urban occlusion conditions of each particular database, as it will be explained later in Section 4.8. Fourth, the first n neighbors of each geodesic extremity are used to fit a straight line defining the prolongation of each curb. In our experiments, n has been set according to the elevation image resolution in order to get 1 m, *e.g.* for an elevation image with k=10 pix/m the number of neighboring pixels is n=10. Fifth, the intersections of these prolongation lines give the position of control points P_{AB} and P_{CB} to fit Bézier curves, as shown in Figure 4.8(b). Finally, quadratic Bézier reconnections R_{AB} and R_{BC} are shown in Figure 4.8(c).

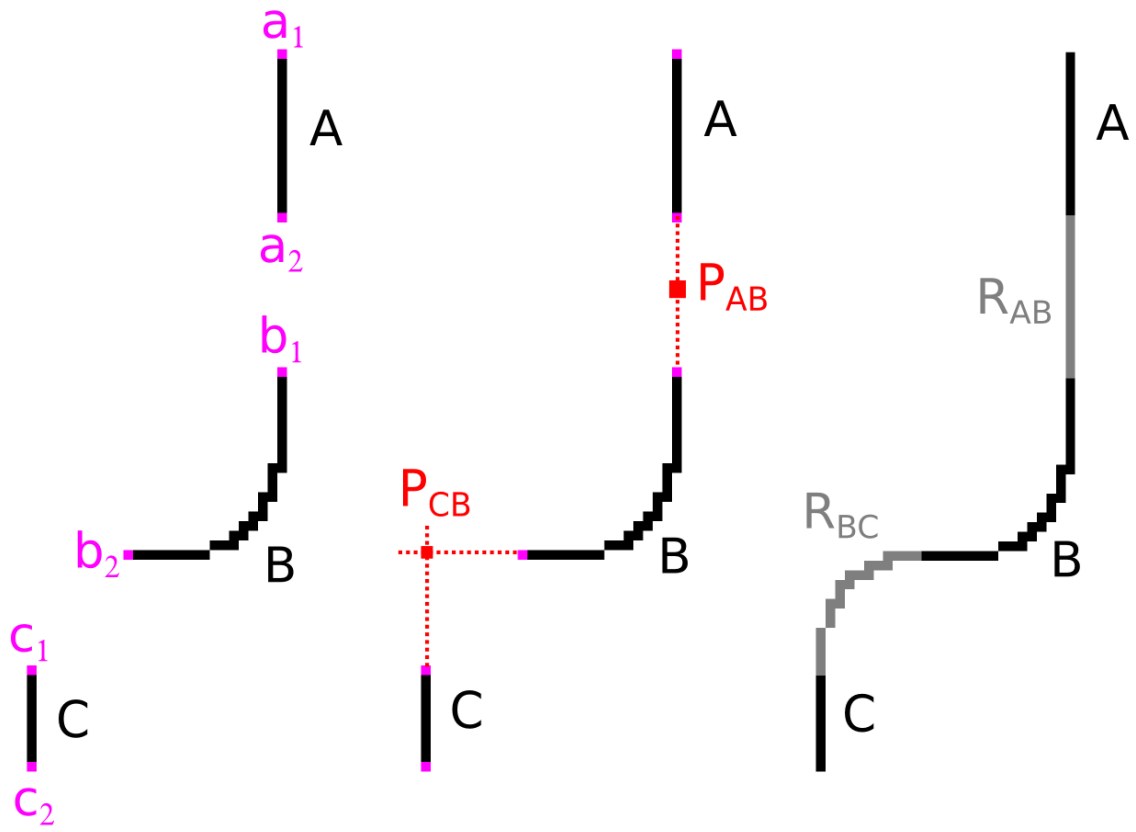


Figure 4.8: Toy example of curb reconnection using Bézier curves. (a) curbs to be reconnected (black) and their geodesic extremities (magenta); (b) control points defined by the intersection of prolongation lines (red); (c) reconnections using quadratic Bézier curves (gray).

Figure 4.9 illustrates this procedure in a real case. Note that the resulting curve is smooth and faithful to reality.

4.5.2 Curb reconnection in special cases

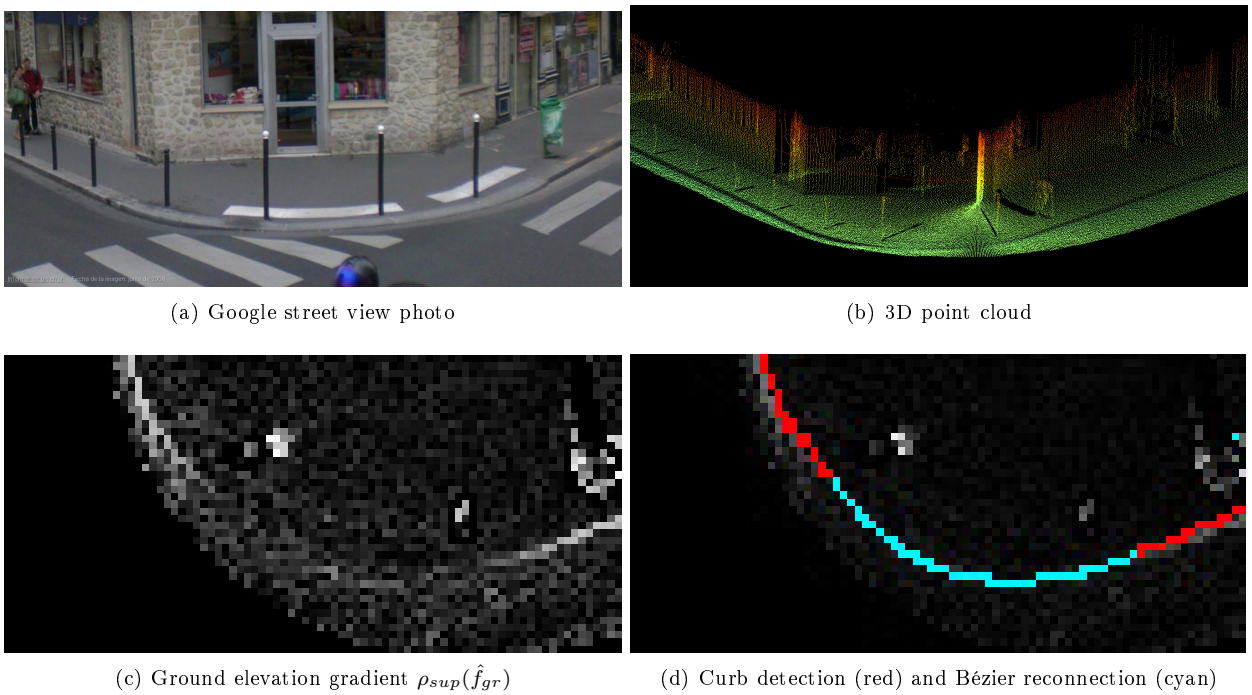
In order to improve curb reconnections and avoid false positives, semantic information has been taken into account in our method. The following special cases are considered:

1. Curbs closer than a given distance d_{facade} from the facade are not used during reconnections because they generally correspond to building entrances. In our experiments, d_{facade} has been set to 40 cm. This parameter is not critical since allowed sidewalks in urban environments should be at least 1.2 m width[2]. After reconnection, these curbs can be reinserted in order to analyze building accessibility.

2. If the Euclidean distance between two curb extremities is greater than d_{min}, no reconnection is allowed. In our experiments, d_{min} has been empirically set to 8 m, which corresponds approximatively to the occluded region produced by two parked cars. This parameter can be adapted according to urban occlusion conditions of each particular database, as it will be explained later in Section 4.8.

3. If Bézier control point P_1 is too distant from the extremities to be reconnected, then P_1 is moved to the

[2] Arrêté du 15 janvier 2007 portant application du décret 2006-1658 du 21 décembre 2006 relatif aux prescriptions techniques pour l'accessibilité de la voirie et des espaces publics. Version consolidée au 03 octobre 2012.



(a) Google street view photo

(b) 3D point cloud

(c) Ground elevation gradient $\rho_{sup}(\hat{f}_{gr})$

(d) Curb detection (red) and Bézier reconnection (cyan)

Figure 4.9: Example of curb reconnection using Bézier curves. Test site in *rue Vaugirard* in Paris, France. Stereopolis II, IGN©.

middle of segment $\overline{P_0P_2}$ and the reconnection becomes straight, as shown in Equation (4.3):

$$\text{if } d(P_0, P_1) > 2d(P_0, P_2) \text{ or } d(P_1, P_2) > 2d(P_0, P_2) \text{ then } P_1 = (P_0 + P_2)/2 \tag{4.3}$$

4. If two curbs are parallel but not co-linear, then there is no intersection between their prolongation lines (mathematically, control point P_1 is at the infinity). In that case, Bézier control point is placed in the middle of segment $\overline{P_0P_2}$ and the reconnection becomes straight, as shown in Equation (4.4):

$$\text{if } P_1 \to \infty \text{ then } P_1 = (P_0 + P_2)/2 \tag{4.4}$$

5. In order to avoid false curb reconnections crossing the road, only reconnections on the same city block are allowed. For this, the city block segmentation method proposed in Section 5.6 is applied. The medial road axes can be used if available. They can be computed directly from the 3D point using the vehicle trajectory information, as shown in Figure 4.11(a), or can be obtained from external 2D maps, as shown in Figure 4.10.

6. One of the most common problems when scanning urban areas is the occlusion due to fix and mobile objects. In particular, parked cars produce large occlusions on the sidewalk, which is a problem during curb segmentation. Figure 4.11 presents an urban test site with several parked cars on the left street side. It is noteworthy that occluded regions restrict curb visibility. In order to solve this problem, semantic information about parked cars, obtained from our object classification method (Chapter 6), is used as follows: the morphological dilation of a parked car over its neighboring occluded region is used as a curb candidate, then the reconnection is allowed reducing the reconnection threshold to $d_{min}/2$. In order to avoid false alarms, if an occluded region is closer than d_{facade} to a facade, to a previously detected curb or to a medial road axis, it is not considered. Figure 4.11(b) shows the result of this reconnection, where detected curbs are marked in red, curbs occluded by parked cars in magenta and curb reconnections in cyan. The result projected onto the 3D point cloud is shown in Figure 4.11(c).

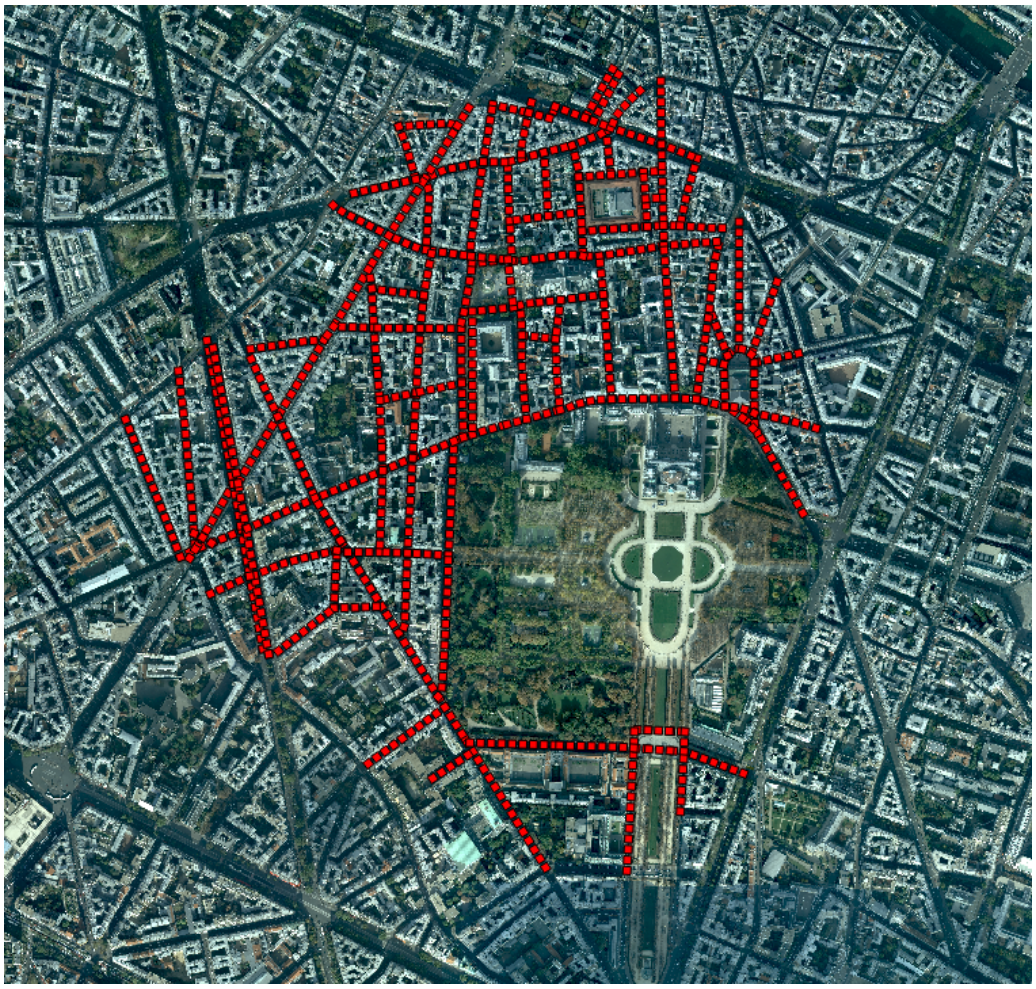
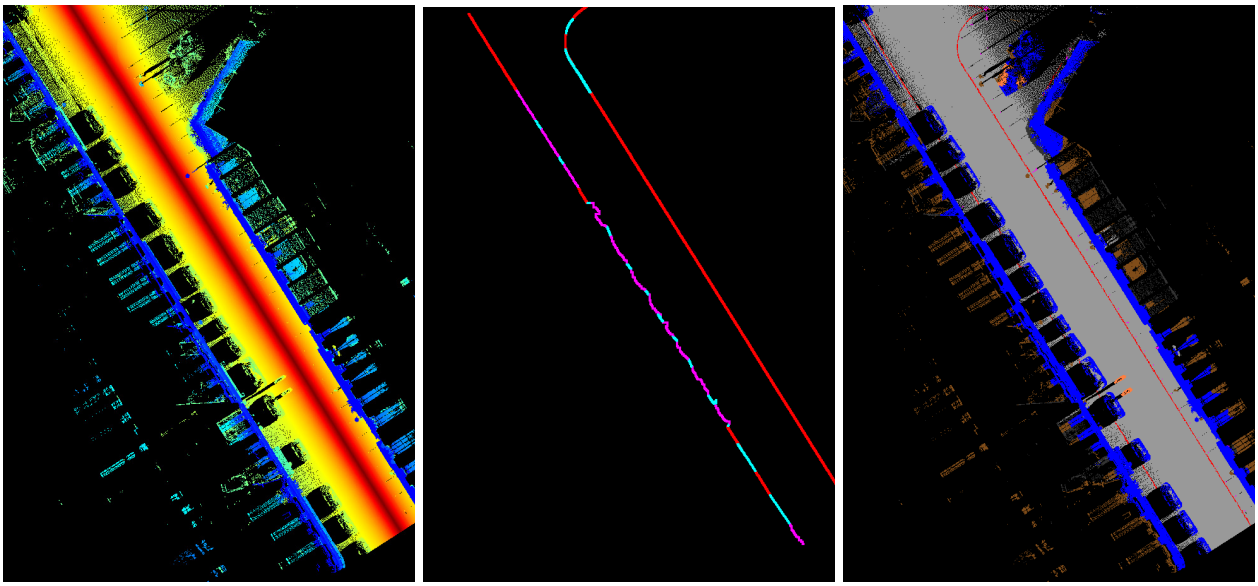


Figure 4.10: Road medial axes in the 6^{th} Parisian district. Information available from IGN France. This information is used in order to avoid false curb reconnections crossing the road, only reconnections on the same city block are allowed.

4.6 Roads and sidewalks segmentation

While curbs define the limit between roads and sidewalks, their explicit segmentation is important because it defines the available zone for vehicles and pedestrians, respectively.

In order to segment roads and sidewalks, a morphological segmentation process based on a constrained watershed is used. The process is shown in Figure 4.12 and is explained as follows: i) from previously segmented and reconnected curbs (Figure 4.12(b)), the distance function is computed. For each pixel, the distance function is defined as the distance from that pixel to the closest curb point. Figure 4.12(c) shows the distance function while Figure 4.12(d) shows its inverse. Note that "no data" pixels on the interpolated elevation image are not considered; ii) facades and road medial axes are used as markers for sidewalk and road, respectively (Figure 4.12(e)); and, iii) a constrained watershed is applied to the inverted distance function. The result is the sidewalk and road segmentation, as shown in Figure 4.12(f). Note that border effects may appear if entire curbs or facades do not appear in the elevation image. In order to solve this problem, overlapping zones are recommended in a large scale application. In order to visualize the segmentation result in the 3D space, a reprojection onto the point cloud is shown in Figure 4.17.



(a) 3D point cloud colored by the angle information.

(b) Curb reconnection using parked cars information. Detected curbs (red), occluded curbs (magenta), curb reconnections (cyan).

(c) Classified point cloud. ground (gray), curbs (red), facades and objects (blue), other (brown).

Figure 4.11: Curb reconnection using semantic information about vehicle trajectory and parked cars. Acquisition by IGN©France. This urban test site presents several parked cars on the left street side. It is noteworthy that occluded regions restrict curb visibility. In order to solve this problem, semantic information about parked cars is used as follows: the morphological dilation of a parked car over its neighboring occluded region is used as a curb candidate, then the reconnection is allowed reducing the reconnection threshold to $d_{min}/2$. In order to avoid false alarms, if an occluded region is closer than d_{facade} to a facade, to a previously detected curb or to a medial road axis, it is not considered.

4.7 Accessibility analysis and itinerary planning

One of the aims of TerraMobilita project is planning itineraries for different types of mobility, including soft-mobility, as presented in Section 1.3. Therefore, curb characterization is a very important task because it determines the suitability of a path. For example, a sidewalk without access ramps may be appropriate for rollers but not for wheelchairs. Additionally, obstacles on the sidewalk represent physical barriers to free mobility. In our work, we define the accessibility according to curbs geometry and obstacles on the street. In our opinion, the most critical case is the accessibility for wheelchair users, so our experiments are conducted in that sense. However, we can define the accessibility according to any other type of soft–mobility since our method provides geometrical information of curbs and obstacles for each point.

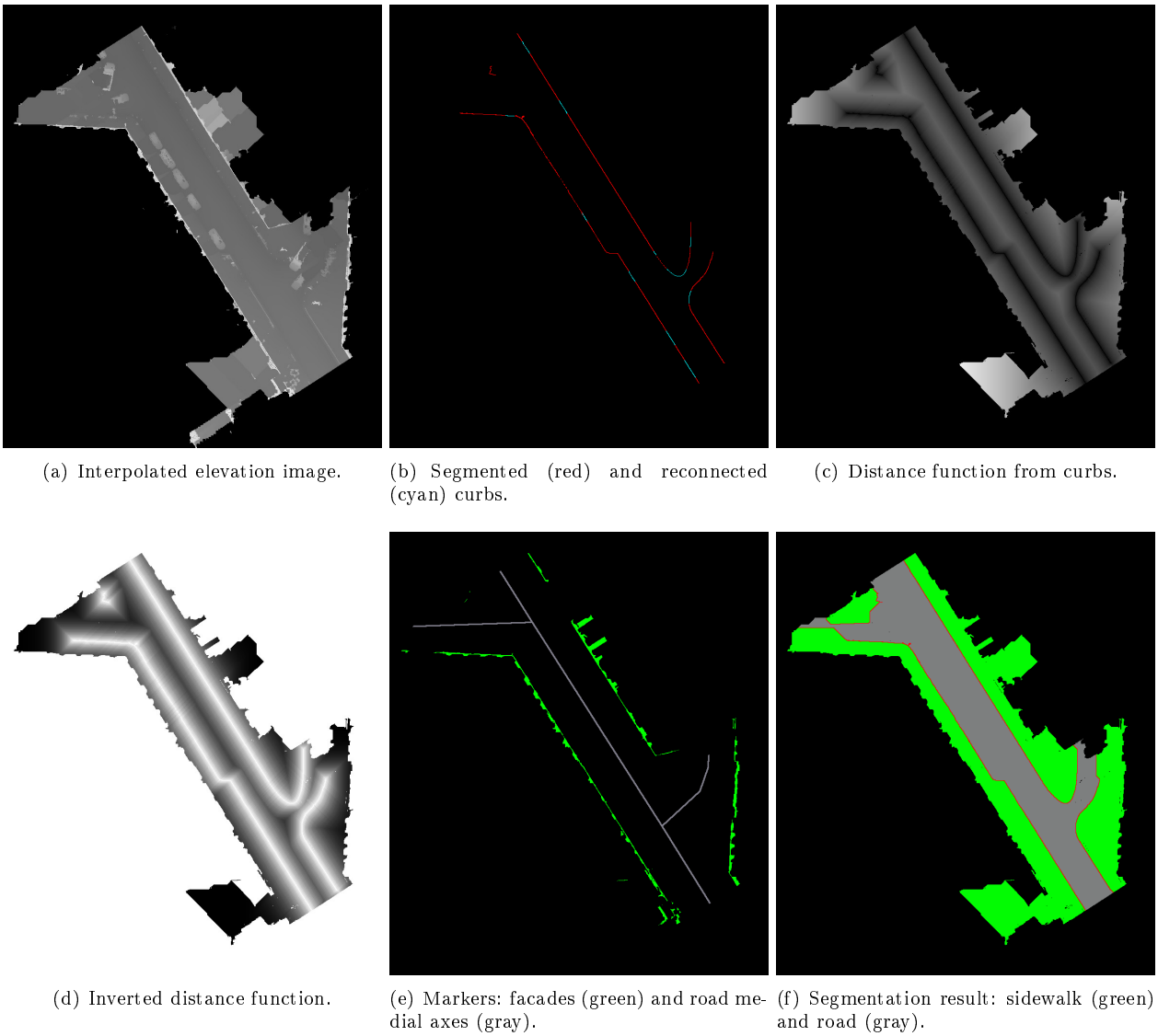
A standard wheelchair is between 60 and 69 cm wide, therefore the minimum clear width of an access ramp is 91.5 cm between railings (ISO, 2008; ADA, 2010). Thus, curb accessibility is defined taking the following criteria into consideration:

Wheelchair–accessible: sidewalk access with one step maximum, wider than 1 m and not higher than 7cm. Wheelchair–accessible curbs are marked in green.

Wheelchair–inaccessible: otherwise. Wheelchair–inaccessible curbs are marked in red.

The simple traffic-light color code (green: accessible, red: non-accessible) is strongly inspired by international standards and is compatible with on–line maps such as Wheelmap (Sozialhelden, 2012). Figure 4.13 illustrates two labeled 3D point clouds from two test sites in *rue Cassette* and *rue Vaugirard* in Paris, France. Note that curbs are correctly segmented and their accessibility defined for a person using a wheelchair.

A direct application consists in planning adaptive itineraries for different types of mobility. For example, defining the start and final points of a journey, it is possible to suggest an adaptive itinerary according to



(a) Interpolated elevation image.

(b) Segmented (red) and reconnected (cyan) curbs.

(c) Distance function from curbs.

(d) Inverted distance function.

(e) Markers: facades (green) and road medial axes (gray).

(f) Segmentation result: sidewalk (green) and road (gray).

Figure 4.12: Roads and sidewalks segmentation using a constrained watershed approach. Acquisition by IGN©France. (a) presents the interpolated elevation image. From segmented and reconnected curbs (b), the distance function is computed (c). (d) shows the inverse distance function. Note that "no data" pixels on the interpolated elevation image are not considered. (e) facades and road medial axes are used as markers for sidewalk and road, respectively. Finally, a constrained watershed is applied to the inverted distance function. The result is the sidewalk and road segmentation, as shown in (f).

obstacles on the ground and curb accessibility. Thus, the problem consists in finding a path that optimizes certain criteria (i.e., the shortest path). Figure 4.14 presents an example of an adaptive itinerary for a person using a wheelchair going from A to B. In this case, we assume that it implies a minimum passing space of 1 m, which is large enough for a standard wheelchair.

Note that this example is only illustrative, real applications for itinerary planning will be developed in the framework of TerraMobilita project. For this, it is necessary to export the obstacle map and the accessibility information into a Geographical Information System (GIS), where adaptive itineraries can be defined according to different types of soft-mobility. Figure 4.15 shows an example of this exporting into a GIS on a test site in *St. Sulpice* square in Paris, France. Note that obstacle position and curb height have been exported as well.

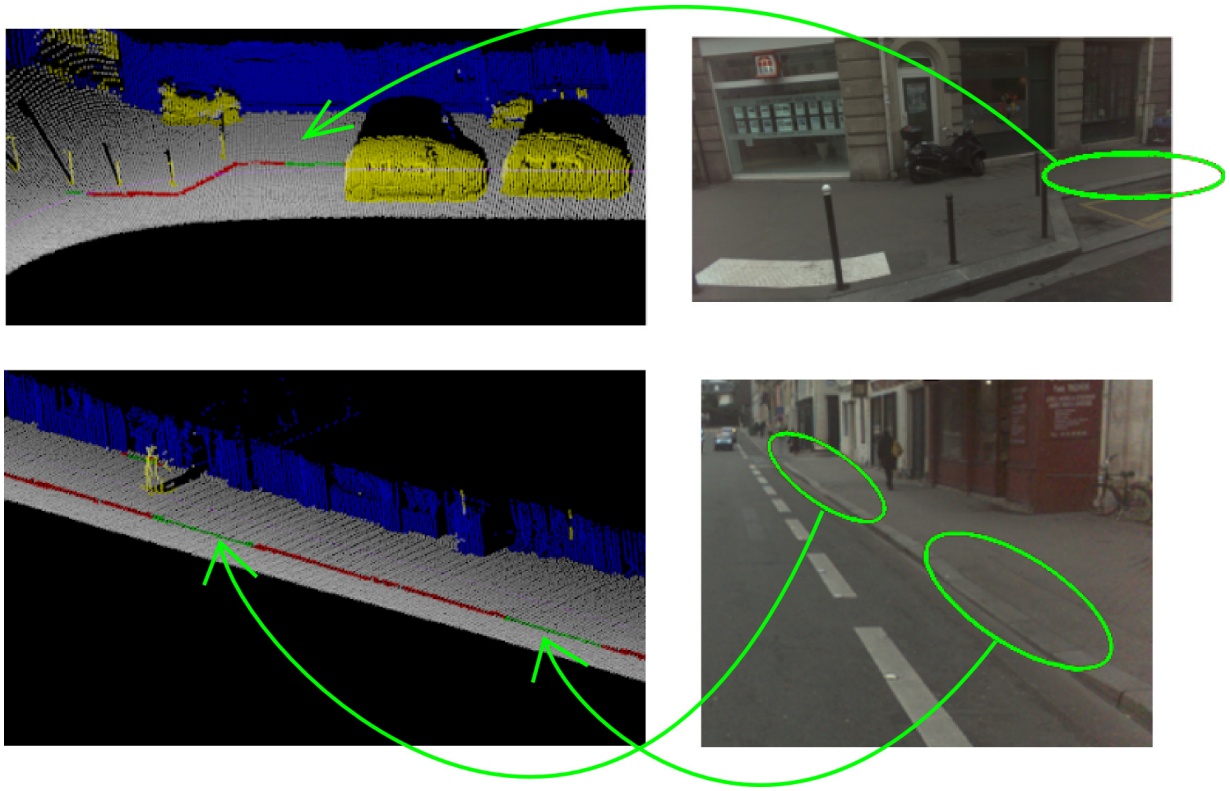


Figure 4.13: Curbs accessibility on two test sites in *rue Cassette* and *rue Vaugirard* in Paris, respectively. 3D labeled point clouds: facades (blue), urban objects (yellow), ground (gray), inaccessible curb (red), accessible curb (green). Acquisition by IGN©France.

4.8 Results

In order to get qualitative and quantitative results, our methodology has been tested on two publicly available databases: TerraMobilita/iQmulus database (Section 2.6.2) and Enschede database (Section 2.5.4). In the first case, our ground and obstacle map segmentation methods are evaluated point by point. In the second case, 2D manual annotations are used to evaluate our curb segmentation and reconnection approach.

Classical Precision (P), Recall (R) and $f_{mean} = (2 \times P \times R)/(P + R)$ statistics are computed. Details are given in the following subsections.

4.8.1 TerraMobilita/iQmulus database

For this experiment, "Cassette_idclass.ply" file has been used[3]. It contains 12 million points from a street section approximatively 200 m long in *rue Cassette* in Paris, France. Manual annotations and point-wise evaluations have been independently carried out by the National French Mapping Agency (IGN).

Figure 4.16(a) presents the input 3D point cloud colored by the laser intensity, while Figure 4.16(b) shows the 3D point cloud processed by our method. Ground appears in gray, facades in blue, and other objects in yellow.

Our evaluation uses the hierarchy of semantic classes defined in TerraMobilita/iQmulus benchmark (Section 2.6.2). First, we classify the 3D point cloud in 3 main categories: *surface* (containing facades and ground), *object* and *other*. Moreover, we define the *unclassified* category for non-annotated points in the GT. They are ambiguous points difficult to annotate, which correspond to 18.31 % of total number of 3D points in the

[3] The manual annotated 3D point cloud is available at:
http://data.ign.fr/benchmarks/UrbanAnalysis/download/Cassette_idclass.zip
The 3D point cloud processed by our method is available at:
https://partage.mines-telecom.fr/public.php?service=files&t=294aed38d48c8ddd03a528069f1b2e51

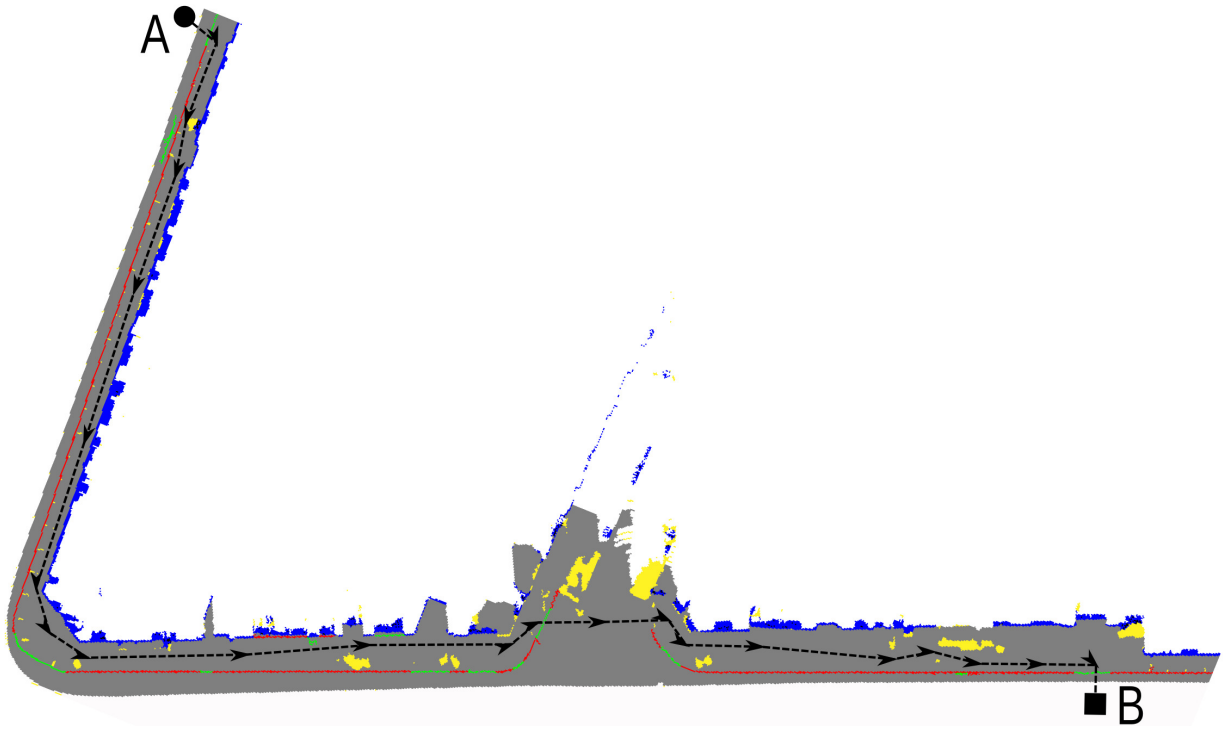


Figure 4.14: Example of an adaptive itinerary (black) for a person using a wheelchair going from A to B. We impose a minimum passing space of 1 m, which is large enough for a standard wheelchair. Nadir view of the scene: facades (blue), urban objects (yellow), ground (gray), inaccessible curb (red), accessible curb (green). Acquisition by IGN©France.

dataset. This classification is important to evaluate our obstacle map generation method. Table 4.1 presents the confusion matrix and our classification results for these 4 categories.

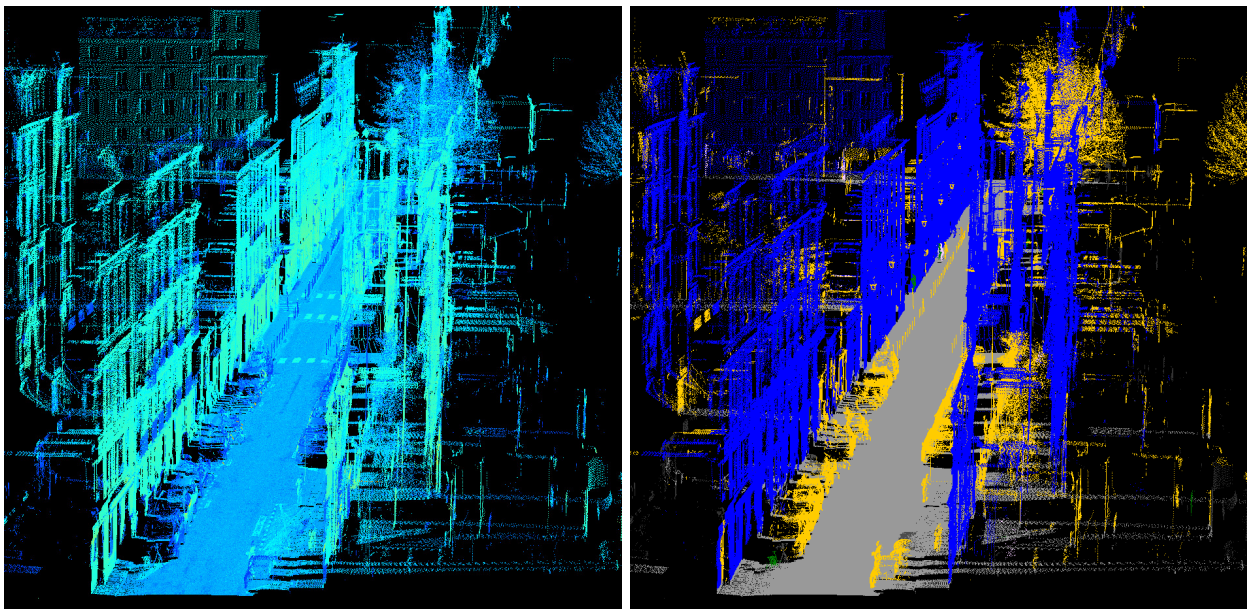
Using our segmentation method, *surfaces* and *objects* are mainly correct with f_{mean} equal to 96.03 % and 84.59 %, respectively. Note that the *surface* class includes facades and ground, which represents the largest structure in the scene with 75.82 % of total number of 3D points, while the *object* class represents 5.7 % of total number of 3D points. Figure 4.17(b) shows a typical segmentation error due to low facades wrongly detected as objects. The *other* class is not correctly classified by our method. However, it is not critical since it only represents 0.17 % of all 3D points in the scene. The *unclassified* class is not critical in the practical case since it mainly contains 3D points behind facades, therefore they do not affect urban mobility. The overall accuracy of our method is 92.65 %.

Table 4.1: Evaluation taking into account 4 main categories on TerraMobilita/iQmulus database. GT: ground truth, AR: Automatic result. In the confusion matrix, results are presented as percentages with respect to the total number of points in the 3D point cloud (12 million points).

GT/AR	**unclassified**	**other**	**surface**	**object**	**Sum**	Recall	Precision	f_{mean}
unclassified	-	-	-	-	18.31 %	-	-	-
other	0.00 %	0.00 %	0.13 %	0.04 %	0.17 %	0.59 %	0.05 %	0.08 %
surface	1.90 %	2.19 %	70.81 %	0.91 %	75.82 %	93.40 %	98.82 %	96.03 %
object	0.09 %	0.02 %	0.72 %	4.88 %	5.70 %	85.49 %	83.72 %	84.59 %
Sum	1.99 %	2.21 %	71.66 %	5.82 %	81.69 %	**Overall accuracy: 92.65 %**		

Table 4.2 presents our segmentation results for the surface class. Note that our method correctly separates facades and ground giving f_{mean} equal to 97.25 % and 98.72 %, respectively. Figure 4.17 shows that small errors are due to the facade-ground junction, where some points may be wrongly assigned. The overall accuracy in this case is 98.26 %. These results prove the performance of our method.

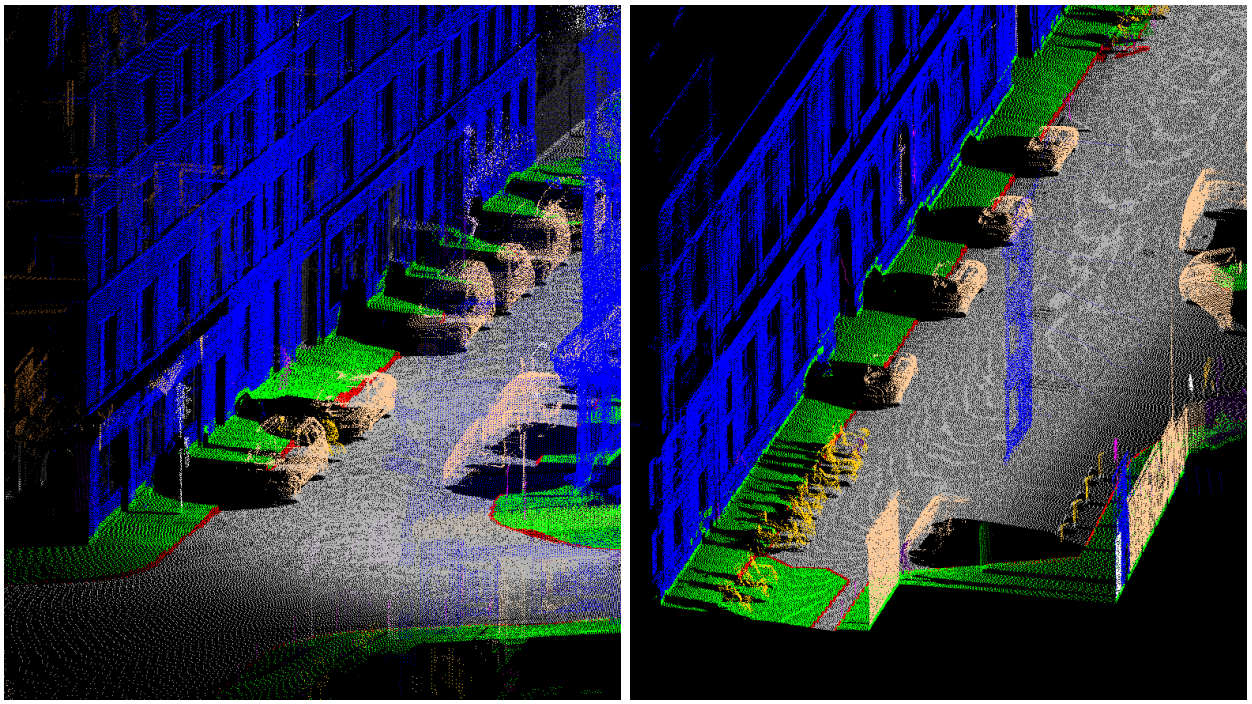
Figure 4.15: Obstacle map and accessibility information exported into a GIS. Curbs (orange), bollards (magenta), medial road axes (red), obstacle bounding box (blue).



(a) 3D point cloud colored by the intensity information.

(b) Classified point cloud. Ground (gray), facades (blue), objects and other (yellow).

Figure 4.16: Ground segmentation and obstacle map definition. Test site in *rue Cassette* in Paris, France. Input file taken from TerraMobilita/iQmulus database. Acquired by Stereopolis II system, IGN©France.



(a) Segmentation errors in the facade–ground junction.

(b) Segmentation errors in the facade–ground junction and a facade in the right part wrongly classified as object.

Figure 4.17: Some errors when segmenting ground, facades and objects. These are typical segmentation errors due to low facades wrongly detected as objects or the lower part of the facade has been wrongly segmented as ground.

Table 4.2: Evaluation taking into account only the surface class (facades and ground) on TerraMobilita/iQmulus database. GT: ground truth, AR: Automatic result. In the confusion matrix, results are presented as percentages with respect to the total number of points in the 3D point cloud (12 million points).

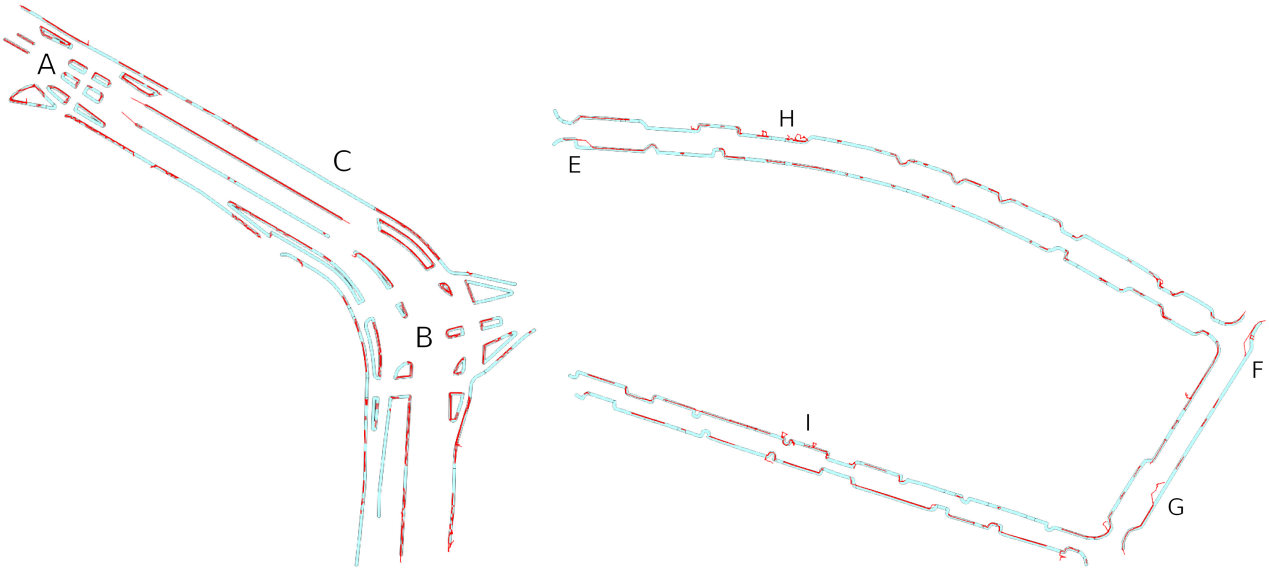
GT/AR	**ground**	**facade**	**Sum**	Recall	Precision	f_{mean}
ground	30.77 %	0.01 %	30.78 %	99.96 %	94.69 %	97.25 %
facade	1.73 %	67.49 %	69.22 %	97.51 %	99.98 %	98.72 %
Sum	32.50 %	67.50 %	100.0 %	**Overall accuracy:**		**98.26 %**

4.8.2 Enschede database

In order to benchmark our curb segmentation and reconnection methods with other state of the art methods, we use another publicly available database containing three test sites in Enschede, The Netherlands (Section 2.5.4). This database has been previously used by Vosselman and Zhou (2009); Zhou and Vosselman (2012); Serna and Marcotegui (2013b), thus comparison with the state of the art becomes possible.

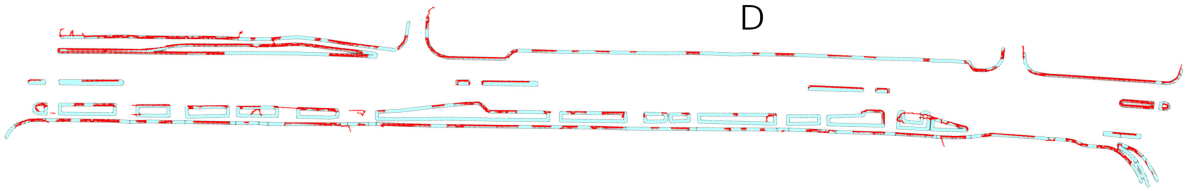
Enschede database contains approximatively 1000 m of MLS data with 2D manual annotations. Two manual ground truth data have been collected: i) roadside lines, corresponding to inaccessible curbs higher than 7 cm; and, ii) gap lines, corresponding to access ramps lower than 7 cm. We use the same evaluation strategy than Vosselman and Zhou (2009); Zhou and Vosselman (2012). Quantitative analysis is performed by comparison between automatic and manual extracted lines. As the amount of false alarms near real road lines is very low in this database, a buffer around ground truth lines is taken. Automatic lines are labeled as true positives or false positives if they are located inside or outside the buffer, respectively. A buffer width of 50 cm is used (the same used by other authors in the same database). Two classical statistics are computed: recall (or completeness), defined as the length of the extracted lines inside the buffer divided by the length of the reference lines; and precision (or correctness), defined as the length of the extracted lines inside the buffer divided by the length of all extracted lines.

Figure 4.18 illustrates our automatic curb segmentation results in the three test sites. Results of precision, recall and processing time are given in Table 4.3. In order to simplify lines geometry, the well–known Douglas and Peucker (1973) algorithm was used with 20 cm distance threshold.



(a) Test site 1

(b) Test site 3



(c) Test site 2

Figure 4.18: Curb segmentation on *Enschede* dataset. Our detection (red) and ground truth (cyan). Enschede database contains approximatively 1000 m of MLS data with 2D manual annotations. Two manual ground truth data have been collected: i) roadside lines, corresponding to inaccessible curbs higher than 7 cm; and, ii) gap lines, corresponding to access ramps lower than 7 cm. Quantitative analysis is performed by comparison between automatic and manual extracted lines. Automatic lines are labeled as true positives or false positives if they are located inside or outside an evaluation buffer, respectively. A buffer width of 50 cm is used (the same used by other authors in the same database).

Our results show that our method has good detection rates, is fast and presents few false alarms. On the one hand, precision is greater than 90% for all sites, which indicates that our method produces few false alarms, mainly due to low vegetation (Zones H and I in Figure 4.18(b)). Moreover, precision is up to 7 % better than other works reported in the literature on the same database. On the other hand, recalls in sites 2 and 3 are better than other works reported in the literature on the same database. However, test site 1 leads to a low recall because of polygonal curbstones in the middle of the road (Zones A and B in Figure 4.18(a)). Since MLS data was acquired only from one side of the street, only one side of the polygons is visible. As aforementioned in Section 4.5, gradients touching an interpolated zone are not considered in order to avoid false alarms. As our original goal consists in detecting curbs delimiting the sidewalk, our method does not process polygonal curbstones in any special way, then the invisible part is not detected while scores published by Zhou and Vosselman (2012) take these polygons into account. Fitting polygons can be a suitable solution and it will be evaluated in our future work. Another problem in site 1 is due to long access ramps that cannot be reconnected by our method. For example, zone C in Figure 4.18(a) shows an access ramp lower than 3 cm and 45 m long. Therefore, it is neither detected nor reconnected.

The presence of cars and other obstacles is the main problem in the detection procedure. In fact, several curbs

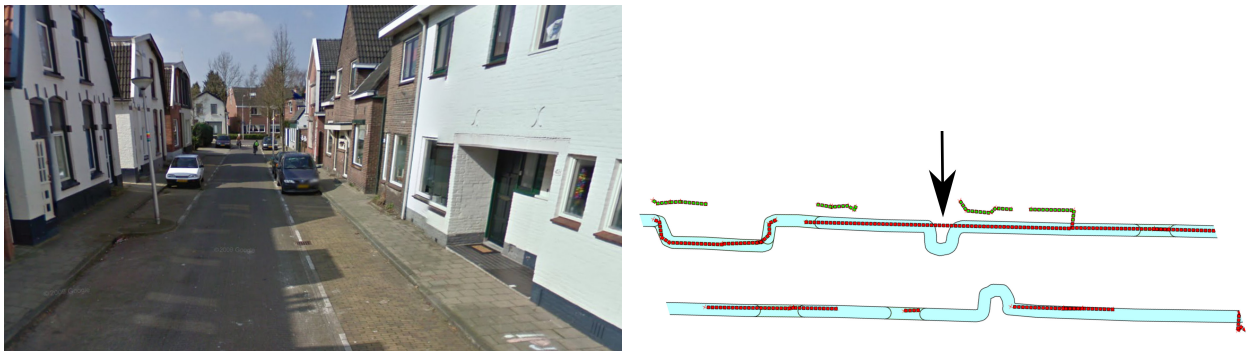
Table 4.3: Precision, recall and processing time for the three test sites at *Enschede*, The Netherlands. Between brackets the results obtained by Zhou and Vosselman (2012)

	Site 1	Site 2	Site 3
Precision	95% (91%)	94% (92%)	91% (84%)
Recall	65% (83%)	54% (53%)	60% (54%)
Time	8.6 min (1 hour)		

are not detected due to large occluded areas. For example, zone D in Figure 4.18(c) shows a large occluded area due to cars on both sides of the street. Only short curb parts are detected between parked cars and they are not reconnected because the distance exceeds our reconnection threshold. For this database, reconnection threshold has been reduced to 2 m due to wrong reconnections in polygonal curbs in the middle of the road. Therefore, curb reconnections longer than 2 m are not allowed and this is the reason of low recall in curbs detection.

Inspecting test sites, we found several inconsistent ground truth lines since they do not correspond to real curbs. For example, Figure 4.19(b) shows a straight detected curb (red) on the right side, while the ground truth (cyan) marks it as an extrusion. A photo from the scene (Figure 4.19(a)) demonstrates that automatic detection is correct in this case. Note that this is a Google Street View photo, taken another day, so parked cars are not the same. Other ground truth problems can be found in zones E, F and G in Figure 4.18(b). In these cases, our method has been penalized in spite of these correct results. However, these annotation errors are rarely found, thus the comparison with the state of the art remains valid.

Note that our method is designed to detect elevation discontinuities on the ground, not only curbstones. Therefore, stairs and steps at building entrances are detected as well (Figure 4.19). These structures are not errors, but they are not marked in ground truth data. Thus, they were not taken into account in the quantitative results in order to do a fair comparison. To automatically filter out these structures, a constraint of minimal distance (40 cm) from the facade was imposed.



(a) (b)

Figure 4.19: Inconsistent ground truth lines on *Enschede* dataset. Ground truth (cyan) presents an extrusion that does not correspond to the real curb. Detected curbs (red) and building entrances (green). Photo taken from Google Street View, taken another day, so parked cars are not the same.

Processing was carried out on an Intel Core i7 CPU @2.93 GHz with 8 GB RAM. Note that our method takes less than 9 minutes to process the three test sites, which is 7 times faster than any other method running on the same database. Vosselman and Zhou (2009) reported 1 hour for the processing time. The conceptual difference consists in the fact that they process the 3D point cloud on a strip by strip basis, while we project all 3D points to an elevation image and we process them as a complete set using digital image processing techniques.

For further analysis, Table 4.4 presents the individual recall results for each curb type. In general, occlusion affects all detection types. The best recall occurs for roadside lines, which are curbs higher than 7 cm. Long curb reconnections are not allowed and it is the reason of low recall in gap detection. The lowest recall is due to polygonal curbstones. As aforementioned, we only detect one side, then a special processing should be performed for this type of curbs.

Table 4.4: Recall for each curb type on *Enschede* database.

	Site 1	Site 2	Site 3
Roadside (Inaccessible)	82%	67%	64%
Gaps (Accessible)	55%	48%	46%
Polygonal curbstones	46%	45%	N/A

4.9 Conclusions

Urban accessibility affects not only disabled persons but also old people, children and pregnant women. In the framework of the *United Nations convention on the rights of persons with disabilities*, local authorities are required to guarantee accessibility to public spaces in order to reduce social exclusion, low employment and limited education of people concerned by accessibility. Thus, it is very important to be able to make large scale accessibility diagnoses in urban environments. In this chapter, we propose an automatic and robust method for urban accessibility diagnoses using semantic analysis methods on 3D point clouds.

Ground segmentation is one of the most important steps in urban semantic analysis since all the urban entities (facades, objects, etc.) are located on it. Contrariwise to classic methods found in the state of the art, our segmentation method takes advantage of the quasi-flat character of the ground. Therefore, the quasi-flat zones labeling algorithm has been used. It allows to segment the ground even in the presence of access ramps, speed humps and other non–flat structures. Once the ground is extracted, all remaining structures are considered as facades and objects. Discrimination between them is important because facades delimit the public space while urban objects define the obstacle map required for itinerary planning.

Our ground segmentation and obstacle map generation methods have been qualitatively and quantitatively tested on *TerraMobilita/iQmulus* database. Our results show that our method has good detection rates and presents few false alarms. The f_{mean} reported for objects and surfaces detection is equal to 76.06 %. The main drawback is that low walls may be wrongly classified as objects. In the case of separation between ground and facades, f_{mean} is equal to 98.26 %, which proves the efficiency of our approaches. These small errors are due to the junction facade-ground, where some points may be wrongly assigned. A possible solution could include the analysis of normal vectors, as proposed by Deschaud (2010), at the cost of increasing computational time. Another drawback of the method is due to the ambiguity when classifying objects behind facades. However, this is not a critical step since those objects represent a small part of the point cloud and they do not affect urban accessibility.

Once the obstacle map is generated, curb segmentation is the next step in the urban analysis since it defines the edge between roads and sidewalks. This segmentation is important because it defines the available zone for vehicles and pedestrians, respectively. Additionally, curb geometry is used to define the accessibility for a given type of mobility. Using the ground segmentation result, gradient information is used in order to detect elevation discontinuities on the ground. Then, curb candidates are selected, close curbs are reconnected using Bézier curves and characterization is carried out based on geometrical features. Finally, accessibility definition is based on international standards.

Our curb segmentation and reconnection methods have been tested on *Enschede* database. Our results show that our methods have good detection rates, are fast and present few false alarms. In fact, precision and recall results outperform other works reported in the literature on the same database. The main drawbacks are due to occlusions, long access ramps and polygonal curbstones. In order to solve these problems, several scans of the same zone (as those produced by velodyne sensors) can reduce the occlusion; using color gradients can be a suitable solution in order to detect low and long access ramps; and, as suggested by other works in the literature, a special strategy for the invisible parts of polygonal curbstones should be developed. Other problems are due to inconsistent ground truth annotations.

Finally, the obstacle map and the curb accessibility information are combined in order to generate adaptive itineraries for different types of urban mobility. In our opinion, the most critical case is the accessibility for wheelchair users, so our experiments have been conducted in that sense. However, we can define the accessibility according to any other type of soft–mobility since our method provides geometrical information of curbs and obstacles for each point.

As perspective, velodyne data and color images can be used in order to distinguish static from mobile obstacles. Moreover, velodyne and riegl sensors may be combined in order to reduce occlusions problems.

TerraMobilita/iQmulus benchmark 2014 is still open[4], thus other authors can submit their results in order to get comparisons with the state of the art. As aforementioned, the evaluation is independently carried out by the National French Mapping Agency (IGN).

[4] TerraMobilita/iQmulus benchmark 2014: http://data.ign.fr/benchmarks/UrbanAnalysis/ [Last accessed: July 23, 2014.]

5 Facade and city block segmentation

5.1 Résumé

Dans ce chapitre, nous présenterons des méthodes de segmentation automatique de façades à partir de données 3D issues d'un scanner mobile. Après une révision de l'état de l'art, nous présenterons une méthode de segmentation basée sur l'extraction de marqueurs de façades ainsi qu'une autre méthode basée sur le calcul de l'élongation géodésique. Ensuite, à partir des résultats de la segmentation de façade, nous exposerons une méthode automatique de segmentation d'îlots de bâtiments. Finalement, nous reporterons des résultats quantitatifs sur des bases de données disponibles dans la littérature.

5.2 Introduction

Building segmentation can be defined as the process of separating buildings from other objects such as natural and artificial ground, vegetation and urban objects. First researches on automatic building extraction began in the 80s. They used aerial imagery and focused on the extraction of high-level 2D and 3D primitives from stereo images. Unfortunately, those methods may fail since linear primitives are difficult to extract and many of them may not correspond to meaningful geometric features.

In computer vision, elevation images were introduced as data structures allowing direct access to 3D geometric features. First elevation images were mostly acquired from small objects and scenes, using active systems. During the 90s, airborne laser scanning (ALS) became widely available so elevation images of huge scenes and cities became possible. As aforementioned in Chapter 2, accuracy and point density have been improved since then and are still constantly improving. More recently, new acquisition systems such as terrestrial (TLS) and mobile laser scanning (MLS) have been developed, adding not only greater geometrical accuracy, but also facade scans, not visible from ALS (Vosselman and Maas, 2010).

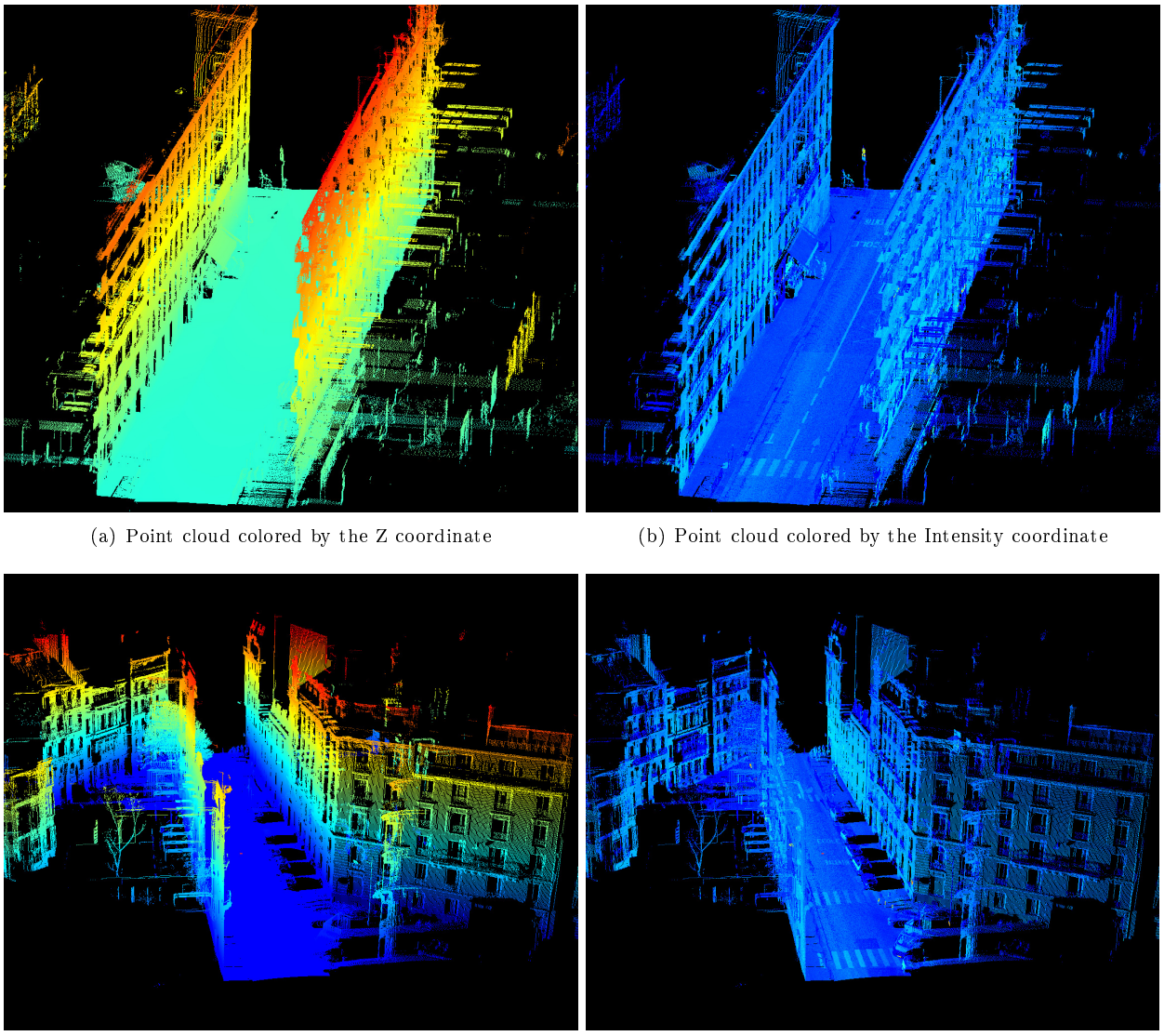
Although the processing of 3D urban data has been underway for many years, facade segmentation is still an open problem. Several contributions on this domain are proposed in the present chapter of this thesis.

Our processing begins with the ground segmentation method proposed in Chapter 4. Once the ground is segmented, all remaining structures are considered as facades and objects. Discrimination between them is important because facades delimit public space and urban objects define the obstacle map required for itinerary planning. Using TLS and MLS data, only building front parts are visible, as shown in Figure 5.1. It is noteworthy that facades constitute the highest and longest vertical entities in the urban scene. In some cases, some artifacts inside buildings may be seen through doors and windows. In this thesis, we are not interested in those objects. In the framework of TerraMobilita project, several facade segmentation methods have been developed according to urban features and application domain.

This chapter is organized as follows. Section 5.3 reviews related works in the state of the art. Sections 5.4 and 5.5 introduce two different approaches to segment facades: with and without markers, respectively. Section 5.6 describes a method to segment city blocks taking advantage of the facade segmentation result. Finally, Section 5.8 concludes this chapter.

5.3 Related work

Goulette et al. (2006a) develop a MLS system, called LARA-3D and used in previous TerraNumerica project (CapDigital, 2009), that acquires and segments in real-time ground, facades and objects. Ground and facades are detected fitting horizontal and vertical planes, then remaining points are considered as objects. In a similar way, Boulaassal et al. (2007) segment building facades using the RANSAC algorithm on TLS data. In general, approaches based on plane extraction are proved to be simple, fast and useful as input for high-level approaches devoted to create accurate geometric models. However, their main drawback is that plane extraction may fail when ground and facades are not flat enough. For example, facades could be over-segmented due to architectural details or balconies.



(a) Point cloud colored by the Z coordinate

(b) Point cloud colored by the Intensity coordinate

(c) Point cloud colored by the Z coordinate

(d) Point cloud colored by the Intensity coordinate

Figure 5.1: 3D point clouds from two test sites in *rue d'Assas* in Paris, France. Stereopolis II, IGN©. Note that facades constitute the highest and longest vertical entities in the urban scene.

Bab-Hadiashar and Gheissari (2006) propose a method to segment planar and curved surfaces in range images. Their method consists in selecting the appropriate parametric model that minimizes strain energy of fitted surfaces. The authors applied their methodology to indoor range images of the University of South Florida (USF) database (Hoover, 1994). Several works on the parametric model fitting problem can be found in the literature (Boyer et al., 1994; Werghi et al., 1998; Marshall et al., 2001; Chaperon and Goulette, 2001; Lari and Habib, 2014). Those works can be extended in order to segment surfaces such as ground and facades on elevation images. The main drawback is that they involve the model selection which can be different for different images, are time consuming due to minimization procedures and may produce under-segmentation.

Demantke et al. (2010) propose a method to adapt 3D neighborhood radius based on local features. Radius selection is carried out optimizing local entropy. Then, dimensionality features are calculated on spherical neighborhoods. These features can be useful to discriminate 1D structures such as pole-like objects, 2D structures such as ground or facades, and 3D volumetric structures such as trees and urban objects.

Hernández and Marcotegui (2009c) assume that facades on the same street are aligned, which is verified in their *Paris-rue-Soufflot* database (Section 2.5.2). They use the Hough transform to detect the facade direction. Then, they analyze the profile of building heights in order to detect facades and city block separations. Hammoudi

(2011) presents a similar technique based on the Progressive Probabilistic Hough Transform in order to detect walls and windows. He assumes that building facades are mainly vertical, so it is possible to generate an accumulation image to compute the number of points projected on the same pixel.

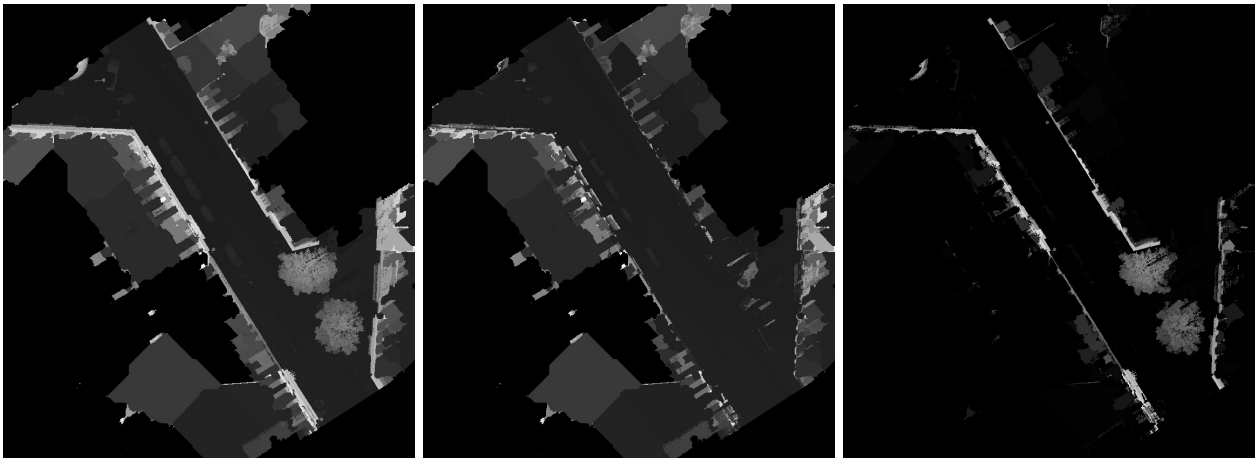
Other works aiming at segmenting facades are available in the literature (Sevcik and Studnicka, 2006; Rutzinger et al., 2011; Poreba and Goulette, 2012b; Serna and Marcotegui, 2013a; Weinmann et al., 2013, 2014). Additionally, facade images can be used in order to enrich the segmentation (Shao et al., 2003; Hernández and Marcotegui, 2009b; Teboul et al., 2010; Serna et al., 2012; Teeravech et al., 2014).

5.4 Facade segmentation using facade markers

In order to segment facades, we propose a method using geometrical constraints in order to define facade markers. Then, a reconstruction is applied from those markers in order to get the entire facade. Let us explain first the facade marker extraction and later in Section 5.4.2 the reconstruction process.

5.4.1 Facade marker extraction

Interpolated relative height image $\hat{f}_{\text{height}}$ is appropriate in order to compute facade markers since it contains information about high and vertical urban structures, as shown in Figure 5.2. Figures 5.2(a) and 5.2(b) show interpolated maximal $\hat{f}$ and interpolated minimal $\hat{f}_{min}$ elevation images, respectively. While Figure 5.2(c) presents the interpolated relative height image, computed as $\hat{f}_{\text{height}} = \hat{f} - \hat{f}_{min}$. For further details on the generation and interpolation of elevation images, the reader is encouraged to refer to Chapter 3 of this thesis.



(a) Interpolated max. elevation image $\hat{f}$ (b) Interpolated min. elevation image $\hat{f}_{min}$ (c) Interpolated relative height image $\hat{f}_{\text{height}} = \hat{f} - \hat{f}_{min}$

Figure 5.2: Interpolated elevation images from a test site in *rue d'Assas* in Paris, France. (a) and (b) show interpolated maximal $\hat{f}$ and interpolated minimal $\hat{f}_{min}$ elevation images, respectively. Note that interpolated relative height image $\hat{f}_{\text{height}}$ is appropriate to compute facade markers since it contains information about high and vertical urban structures, as shown in (c). Acquired by Stereopolis II, IGN©.

From the work by Hernández and Marcotegui (2009a), we reuse the two following geometric constraints on $\hat{f}_{\text{height}}$ in order to extract facade markers:

- *heightFacade*, defining the minimal allowed facade height. In our experiments, this variable has been set to 3.5 m according to architectural characteristics of our databases. This threshold is illustrated in Figure 5.3(a). Note that only the highest objects are preserved.

- *lengthFacade*, defining the minimal allowed facade length. In our experiments, this variable has been set to 5 m according to architectural characteristics of our databases. This threshold is illustrated in

Figure 5.3(b). Note that small objects such as lampposts and objects behind facades are not long enough and are then eliminated.

In addition, we propose a third constraint in order to eliminate round objects such as trees.

- *circularityFacade*, defining the maximal allowed facade circularity (circularity of an object X is defined as the inverse of its elongation $Circ(X) = 1/E(X)$). In our experiments, this variable has been heuristically set to 1/3, which correspond to the circularity of an ellipse whose major axis is 12 times longer than the minor one. Remember that the circularity of a perfect circle is equal to 1. This threshold is illustrated in Figure 5.3(c). Note that non-elongated objects such as the two trees in the right street side have been eliminated.

Using these constraints, we extract facade markers as the union of connected components (CC) higher than *heightFacade*, longer than *lengthFacade* and less circular than *circularityFacade*, as established in Definition 5.4.1:

Definition 5.4.1 *Let $\hat{f}_{height}$ be an interpolated relative height image $\hat{f}_{height}$: $D \rightarrow V$, with $D \subset Z^2$ the image domain and $V = [0, ..., H]$ the set of gray levels mapping the pixel height. Let $Th(\hat{f}_{height})$ be the binary image containing the pixels higher than heightFacade:*

$$Th(\hat{f}_{height}) = \{p \in D | \hat{f}_{height}(p) > heightFacade\} \tag{5.1}$$

Let C_1, C_2,..., C_n be the connected components of image $Th(\hat{f}_{height})$:

$$Th(\hat{f}_{height}) = \bigcup_{i=1}^{n} C_i,\ i \neq j \Rightarrow C_i \cap C_j = \emptyset \tag{5.2}$$

Then, facade markers Fmark of $\hat{f}_{height}$ are the connected components C_i of image $Th(\hat{f}_{height})$ which are longer than lengthFacade and less circular than circularityFacade:

$$Fmark(Th(\hat{f}_{height})) = \{C_j | L(C_j) > lengthFacade \wedge Circ(C_j) < circularityFacade\}; \forall j \in \{1, ..., n\} \tag{5.3}$$

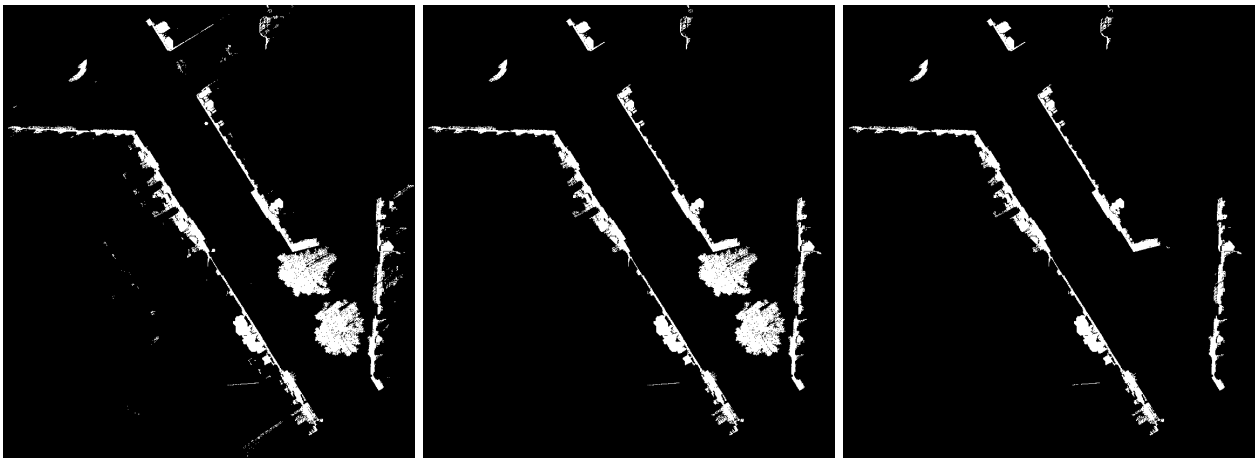
where $L(C_j)$ and $Circ(C_j)$ are respectively the geodesic diameter and the circularity of connected component C_j. For further details on the geodesic elongation, the reader is encouraged to read the Section 7.3.3 of this thesis.

It is noteworthy that these three parameters (*heightFacade*, *lengthFacade* and *circularityFacade*) are easy to tune since they have a physical meaning and depend on urban/architectural constraints. Figure 5.3 illustrates this marker selection process.

Due to specific requirements in some TerraMobilita datasets, several 3D point clouds were acquired with the laser system oriented to the ground. Therefore, structures higher than 2.5 m are out of the laser field of view, as shown in Figure 5.4. This is a challenge for methods using height constraints since high wall parts are not visible.

To solve this problem, we propose a solution taking advantage of the acquisition cycle of the MLS sensor, as shown in Figure 5.5. In our configuration, the sensor spins scanning vertical lines starting from the top. Thus, the first and the last point of each spin correspond to the highest point on the right and on the left street side, respectively. These highest points are usually located on the facade. Selecting these points is automatically carried out detecting sign changes in the *angle of depression* (computed using the sensor position) between consecutive points. Then, these markers are added to image $\mathrm{Th}(\hat{f}_{\mathrm{height}})$ and isolated points are filtered out using the same process as before: only markers longer than *lengthFacade* and less circular than *circularityFacade* are considered as facade markers.

Figure 5.6 illustrates facade marker extraction when the laser sensor is oriented to the ground. Figure 5.6(b) presents the facade markers reprojected onto the 3D point cloud. The test site corresponds to a street section in *rue Vaugirard* in Paris, France.

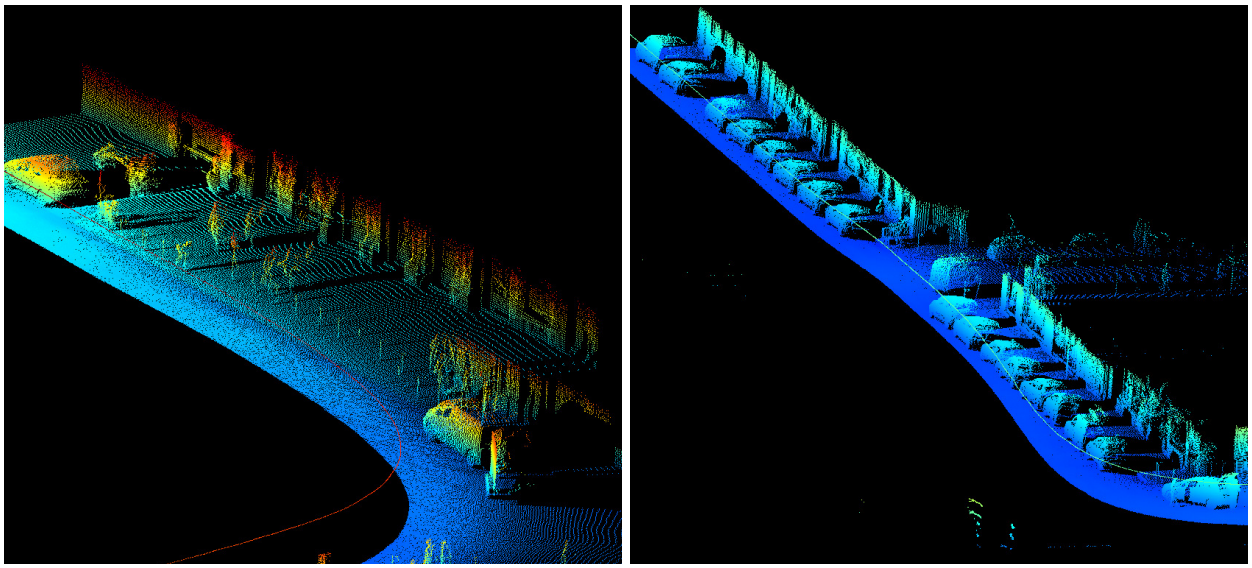


(a) Th($\hat{f}_{\text{height}}$): points higher than *heightFacade*.

(b) Points higher than *heightFacade* and longer than *lengthFacade*.

(c) Facade markers $Fmark(Th(\hat{f}_{\text{height}}))$: points higher than *heightFacade*, longer than *lengthFacade* and less circular than *circularityFacade*.

Figure 5.3: Facade marker extraction using geometrical constraints. It is noteworthy that these three parameters (*heightFacade*, *lengthFacade* and *circularityFacade*) are easy to tune since they have a physical meaning and depend on urban/architectural constraints. Test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©.



(a) 3D point cloud colored by the Z coordinate.

(b) 3D point cloud colored by the Z coordinate.

Figure 5.4: Example of 3D point clouds when laser is oriented to the ground. Therefore, structures higher than 2.5 m are out of the laser field of view. This is a challenge for methods using height constraints since high wall parts are not visible. Test sites in *rue Vaugirard* in Paris, France. Stereopolis II, IGN©.

5.4.2 Facade reconstruction from markers

As aforementioned, facade markers only contain a facade part. Therefore, a reconstruction should be applied from those markers in order to retrieve the whole facade. With this purpose, we use a reconstruction constrained by the ground residue (our ground segmentation method has been previously presented in Section 4.4). Ground residue $\hat{f}^c_{gr}$ is computed as the difference between the elevation image and the ground: $\hat{f}^c_{gr} = \hat{f} - \hat{f}_{gr}$. Then,

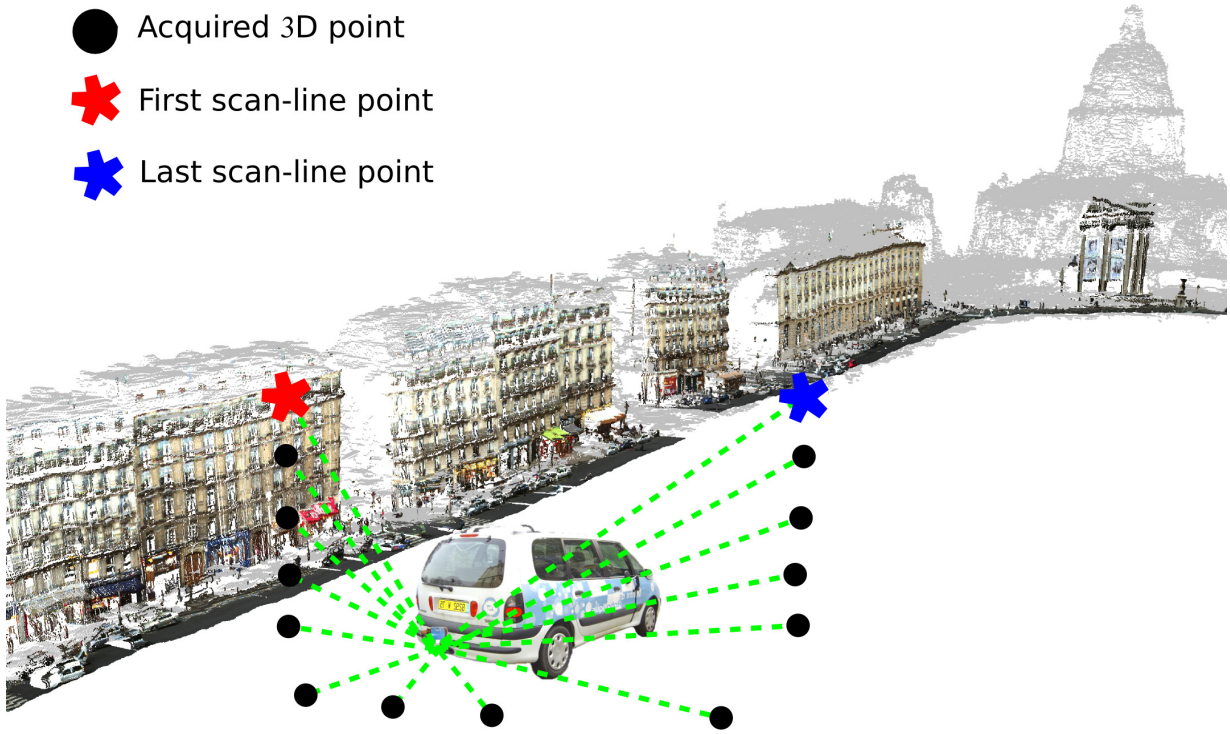
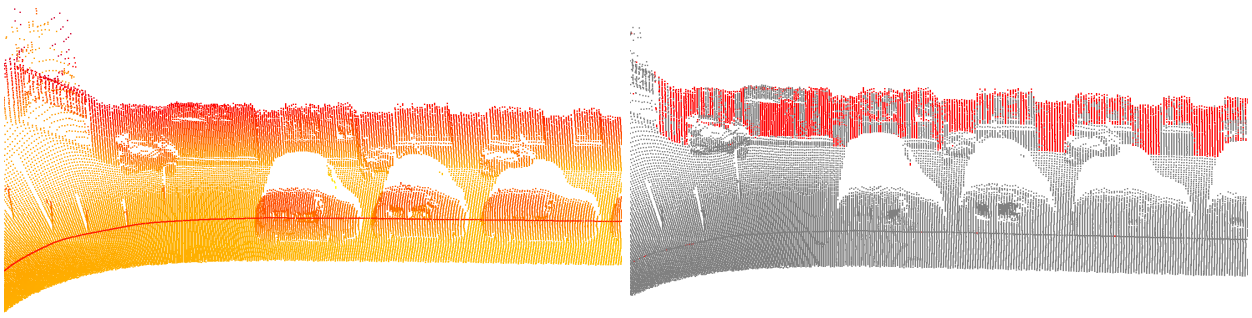


Figure 5.5: Mobile laser scanning (MLS) acquisition cycle. The first and the last point are taken as facade markers. Test site in *rue Soufflot* in Paris, France. LARA-3D, CAOR-MINES ParisTech©.



(a) 3D point cloud colored by the Z coordinate.

(b) Facade markers (red), other (gray).

Figure 5.6: Facade markers reprojected onto the 3D point cloud. During this acquisition, the laser sensor was oriented to the ground. Therefore, structures higher than 2.5 m are out of the laser field of view. This is a challenge for methods using height constraints since high wall parts are not visible. Test site in *rue Vaugirard* in Paris, France. Stereopolis II, IGN©.

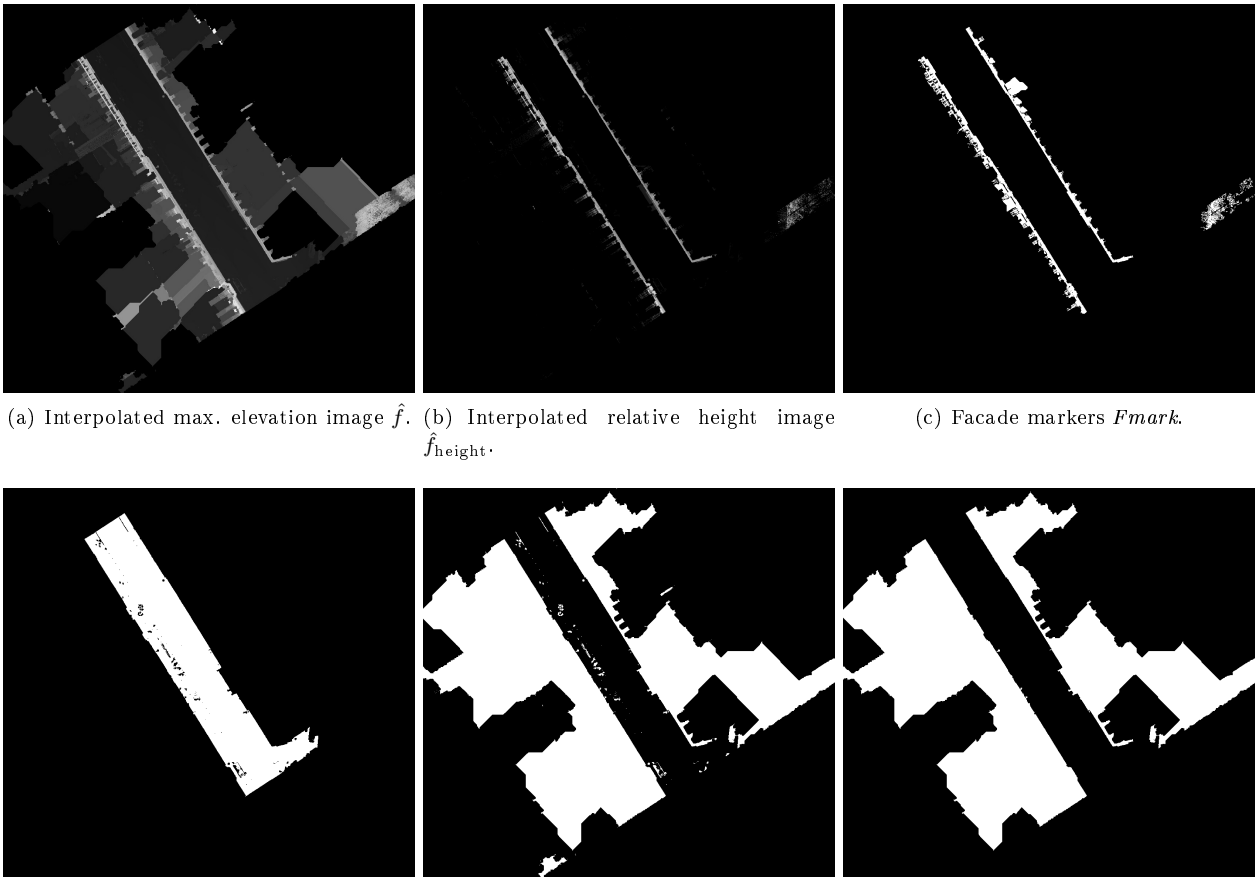
a first solution for the reconstruction process consists in a set of increasing geodesic dilations applied until idempotence. This transformation is called reconstruction by dilation and it is defined as follows:

Definition 5.4.2 ***Reconstruction by dilation.*** *The reconstruction by dilation of a mask image $\hat{f}^c_{gr}$ from a marker $Fmark \leq \hat{f}^c_{gr}$ is defined as the geodesic dilation of $Fmark$ with respect to $\hat{f}^c_{gr}$ until idempotence and it is denoted by:*

$$R^{\delta}_{\hat{f}^c_{gr}}(Fmark) = \delta^{(i)}_{\hat{f}^c_{gr}}(Fmark) \tag{5.4}$$

where i is the geodesic dilation size for which idempotence is reached, i.e. $\delta^{(i+1)}_{\hat{f}^c_{gr}}(Fmark) = \delta^{(i)}_{\hat{f}^c_{gr}}(Fmark)$

Figure 5.7 illustrates the facade segmentation based on reconstruction by dilation. Figure 5.7(a) presents the interpolated elevation image and Figure 5.7(b) the interpolated relative height image. Figure 5.7(c) shows the facade markers *Fmark* computed by the method explained above in Section 5.4.1. Figure 5.7(d) presents the ground segmentation result, while Figure 5.7(e) presents the ground residue $\hat{f}^c_{gr}$. Finally, Figure 5.7(f) shows the facade segmentation obtained by reconstruction by dilation of the ground residue from the facade markers. Note that pixels behind facades have been included in the segmentation result. This method is fast and easy to implement. However, the main drawback is that objects connected to the facade, *e.g.* motorcycles parked next to the facade or pedestrians leaning on walls, are reconstructed as well.



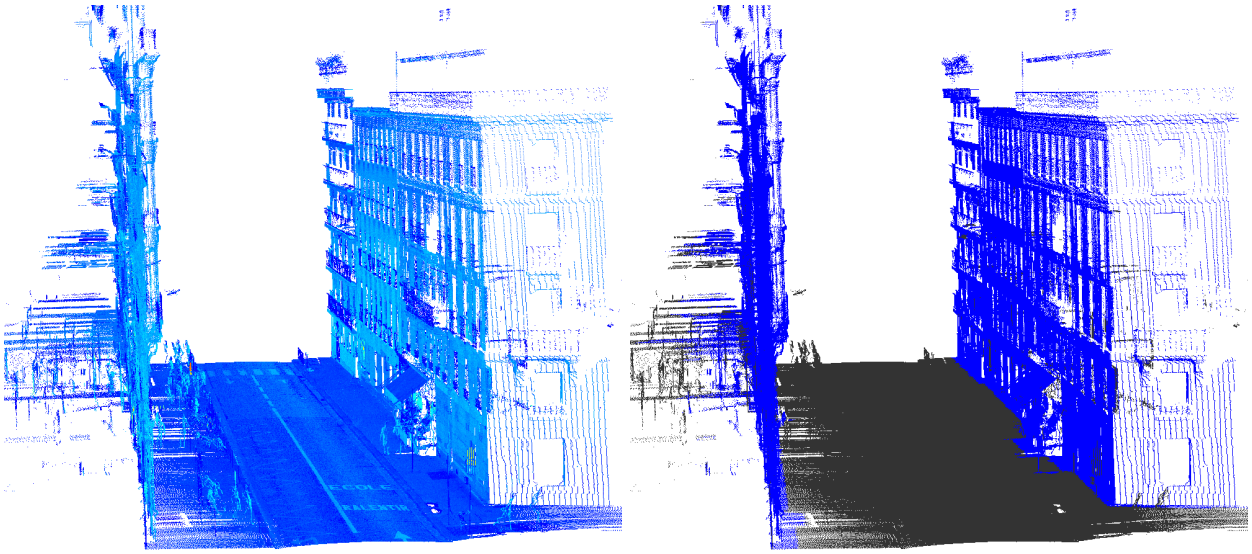
(a) Interpolated max. elevation image $\hat{f}$. (b) Interpolated relative height image $\hat{f}_{\text{height}}$. (c) Facade markers *Fmark*.

(d) Ground segmentation $\hat{f}_{gr}$. (e) Ground residue $\hat{f}^c_{gr} = \hat{f} - \hat{f}_{gr}$. (f) Facade segmentation using reconstruction by dilation: $R^{\delta}_{\hat{f}^c_{gr}}(Fmark)$.

Figure 5.7: Facade segmentation using reconstruction by dilation on the ground residue image from facade markers. Note that pixels behind facades have been included in the segmentation result. This method is fast and easy to implement. However, the main drawback is that objects connected to the facade, *e.g.* motorcycles parked next to the facade or pedestrians leaning on walls, are reconstructed as well. Test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©.

Figure 5.8 presents this segmentation result reprojected onto the 3D point cloud. Note that the reconstruction by dilation retrieves not only the entire facade but also other objects connected to it. For example, the store furniture, a pedestrian and a traffic sign on the right sidewalk have been reconstructed as well.

In order to solve the problem of objects touching the facade, we propose an attribute controlled reconstruction from facade markers. This procedure has been previously published in Serna and Marcotegui (2013a) and detailed in Section 7.4. It consists in appending nearby points with similar height until the maximization of an attribute on the segmented region. In our case, we use increasing propagations from facade markers over



(a) Point cloud colored by the laser intensity.

(b) Segmentation result: facade (blue), other (gray).

Figure 5.8: Facade segmentation result reprojected onto the 3D point cloud. Facade has been segmented using reconstruction by dilation on the ground residue image from facade markers. Note that this reconstruction retrieves not only the entire facade but also other objects connected to it. For example, the store furniture, a pedestrian and a traffic sign on the right sidewalk have been reconstructed as well. Test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©.

quasi-flat zones. As facades are the longest and most elongated structures in the elevation image, we keep the propagation that maximizes the geodesic elongation.

Let us introduce a formal definition for the facade segmentation process using attribute controlled reconstruction:

Definition 5.4.3 ***Attribute controlled reconstruction.*** *Let $\hat{f}^c_{gr}$ be a digital elevation image containing the ground residue $\hat{f}^c_{gr} : D \to V$, with $D \subset Z^2$ the image domain and $V = [0, ..., R]$ the set of gray levels mapping the elevation values. Two neighboring pixels p, q belong to the same λ–flat zone of $\hat{f}^c_{gr}$, if their difference $\left|\hat{f}^c_{gr}(p) - \hat{f}^c_{gr}(q)\right|$ is smaller than or equal to a given λ value.*

For all $x \in Fmark \subseteq D$, let Λ be the set of increasing regions containing marker pixel x. For all $\lambda \in V$ and $j = [1, ..., n-1]$, we define $A_\lambda(Fmark) \in \Lambda$ as the λ–flat zone of image $\hat{f}^c_{gr}$ containing marker Fmark:

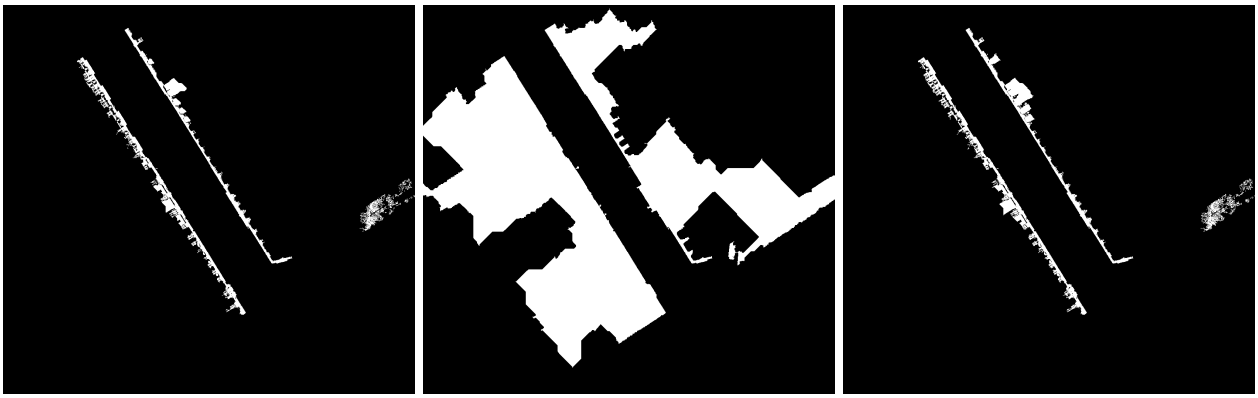
$$A_\lambda(Fmark) = \{x\} \cup \{q | \exists \wp = (p_1 = x, ..., p_n = q) \text{ such that } |\hat{f}^c_{gr}(p_j) - \hat{f}^c_{gr}(p_{j+1})| \leq \lambda\}; \quad \forall x \in Fmark \tag{5.5}$$

Let $E(A_\lambda(Fmark))$ be the geodesic elongation of λ–flat zone $A_\lambda(Fmark)$. For all $\lambda_i \in V$ and $i = [0, ..., R]$, we define λ_M as the value for which the elongation is maximum:

$$\lambda_M = argmax_{\lambda_i \in V} |E(A_{\lambda_i}(Fmark))| \tag{5.6}$$

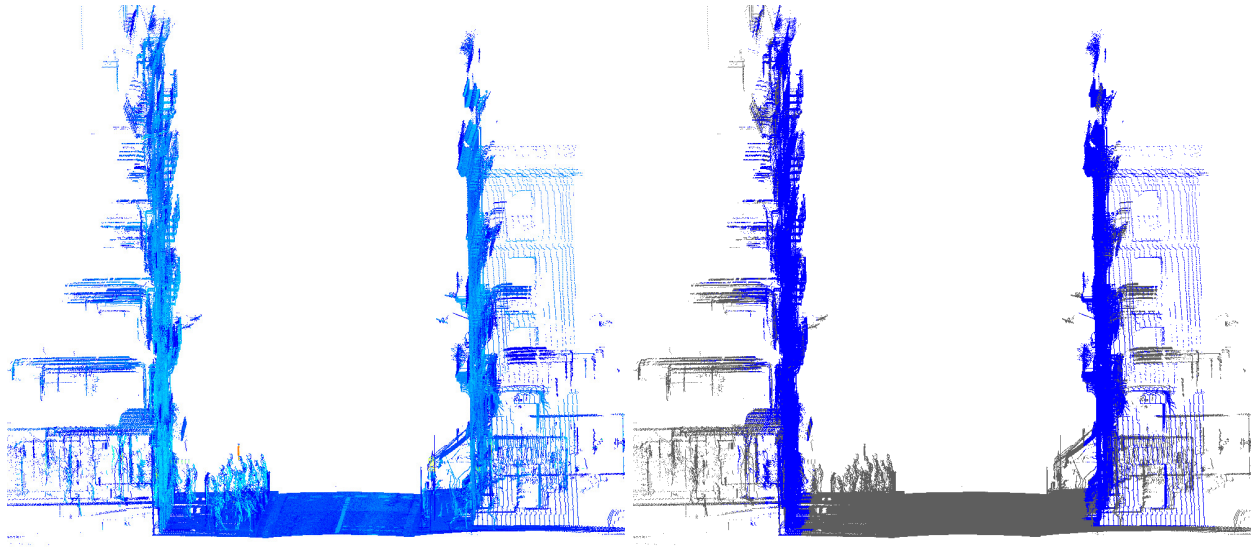
Then, we define $A_{\lambda_M}(Fmark)$ as the attribute controlled reconstruction of the facade from marker Fmark.

Using this controlled reconstruction maximizing the geodesic elongation, it is possible to reconstruct the facade without merging adjacent objects. Figure 5.9 compares facade segmentation methods using reconstruction by dilation $R^\delta_{\hat{f}^c_{gr}}(Fmark)$ and attribute controlled reconstruction $A_{\lambda_M}(Fmark)$ from marker *Fmark*. It is noteworthy that the attribute controlled reconstruction does not reach objects connected to the facade nor objects behind them. Figure 5.10 presents the segmentation result using attribute controlled reconstruction reprojected onto the 3D point cloud. Compared to Figure 5.8, note that the store, the pedestrians and the traffic sign have been correctly separated from the facade. Additionally, several objects such as wall lamps and objects behind facades have been correctly separated.



(a) Facade markers *Fmark*. (b) Reconstruction by dilation $R^{\delta}_{f_{gr}^{c}}(Fmark)$. (c) Attribute controlled reconstruction $A_{\lambda_M}(Fmark)$.

Figure 5.9: Comparison of facade segmentation methods using reconstruction by dilation and attribute controlled reconstruction on the ground residue image. Test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©.



(a) Point cloud colored by the laser intensity. (b) Segmentation result: facade (blue), other (gray).

Figure 5.10: Facade segmentation using attribute controlled reconstruction on the ground residue image. Note that using this attribute controlled reconstruction, the store, the pedestrians and the traffic sign have been correctly separated from the facade. Test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©.

5.4.3 Discussion

Methods based on markers are robust to segment non-planar facades and facades with several architectural details and balconies, such as it is the case of Parisian buildings. Moreover, these methods are simple and fast since they are based on geometric constraints, easily translated into threshold values. In the case of low facades or when the laser sensor is oriented to the ground, additional markers corresponding to the highest points of each laser rotation are added. It is important to remind that markers only contain partial facades, therefore a reconstruction from markers is required in order to get the entire facade. In our case, the reconstruction is constrained to the ground residue.

First, we have proposed a reconstruction by dilation. The main problem is that objects touching the facade, such as motorcycles parked next to the facade or pedestrians leaning on a wall, are included in the segmentation result.

In order to solve this problem, we have proposed a more sophisticated solution using an attribute controlled reconstruction. Since facades appear as high and elongated vertical structures, the choice of the geodesic elongation is justified and very efficient in practice. Additionally, objects touching the facade usually fatten them. Thus, we keep the quasi-flat propagation that maximizes the facade elongation. This method offers better results than the reconstruction by dilation, however it is slower. One solution to speed up the attribute controlled propagation consists in only considering a subset of all possible quasi-flat zones propagations: $\lambda_i = i \times \Delta\lambda;\ \forall i = [0, ..., int(R/\Delta\lambda)]$. In our experiments, we have set $\Delta\lambda$ to 1.0 m since it offers a trade off between processing time and performance, allowing proper separation of connected objects such as parked motorcycles and leaning pedestrians.

The performance of methods based on markers strongly depends on the markers selection. A wrongly located marker may produce errors since it will reconstruct the corresponding object, even if it is not a facade. In our experiments, our marker selection method has proved to be efficient in many cases. However, objects such as tree alignments may produce false markers and then wrong segmentations, as shown in Figure 5.11. That is why we have proposed a more robust method without facade markers. Using such method, only the elongation and its evolution over the height decomposition of the scene are analyzed. This method is proved to produce the best results, as explained in the following section.

5.5 Facade segmentation without markers

In order to segment facades avoiding the use of markers, we propose a method based on threshold decomposition and attribute profiles. This method will be revisited later in Section 7.3.2, where we propose a method to segment elongated objects on gray-scale images. Let us present its definition in the 2D case:

Definition 5.5.1 *Let I be a digital gray-scale image $I : D \rightarrow V$, with $D \subset Z^2$ the image domain and $V = [0, ..., R]$ the set of gray levels. A decomposition of I can be obtained considering successive thresholds:*

$$T_t(I) = \{p \in D | I(p) > t\}\ \forall t = [0, ..., R-1] \tag{5.7}$$

Since this decomposition satisfies the inclusion property $T_t(I) \subseteq T_{t-1}(I), \forall t \in [1, ..., R-1]$, it is possible to build a tree, called the component tree, with level sets $T_t(I)$. Each branch of the tree represents the evolution of a single connected component X_t. An attribute profile is the evolution of an attribute (e.g. area, perimeter, elongation, average gray–level, etc.) of the CC along a branch of the tree.

Figure 5.12 illustrates the threshold decomposition for a 1D function, its component tree and the attribute (width) profiles for the two function maxima (p_A and p_B). Events on this attribute profile are useful to segment objects (Jones, 1999), extract features (Pesaresi and Benediktsson, 2001; Beucher, 2007; Morard et al., 2011b) and define adaptive structuring elements (Serna and Marcotegui, 2013a).

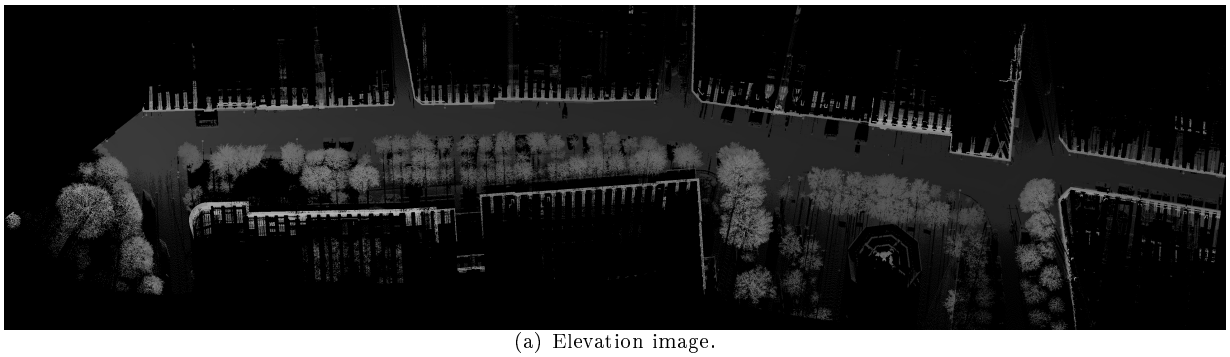
Now, let us extend this definition to the 3D case:

Definition 5.5.2 *Let P be a 3D point cloud containing a list of N points $(x_i, y_i, z_i) \subset \Re^3$, where $i = [0, ..., N]$. Let $M_i = (x_i, y_i, z_i)$ be a 3D point in $\Re^3$ and $m_i = (u_i, v_i)$ a point in Z^2 in elevation image f, resulting from the image projection process explained in Section 3.4.1. A decomposition of P in horizontal slices can be obtained considering successive thresholds on the Z axis separated by a given height Δz:*

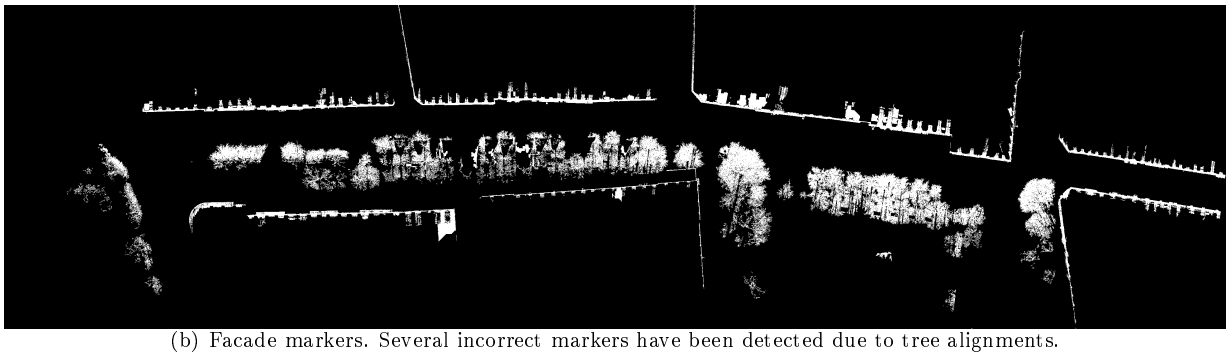
$$T_t^Z(P) = \{m_i \in D \mid t\Delta z < z_i < (t+1)\Delta z\};\ \forall t = [0, ..., R-1];\ \forall i = [0, ..., N] \tag{5.8}$$

Contrarily to the 2D case, this decomposition does not satisfy any inclusion property. However, it is always possible to analyze the evolution of a single connected component X_t over horizontal slices $T_t^Z(P)$. An attribute profile is the evolution of an attribute (e.g. number of points, density, average elevation, etc.) of a CC along the decomposition.

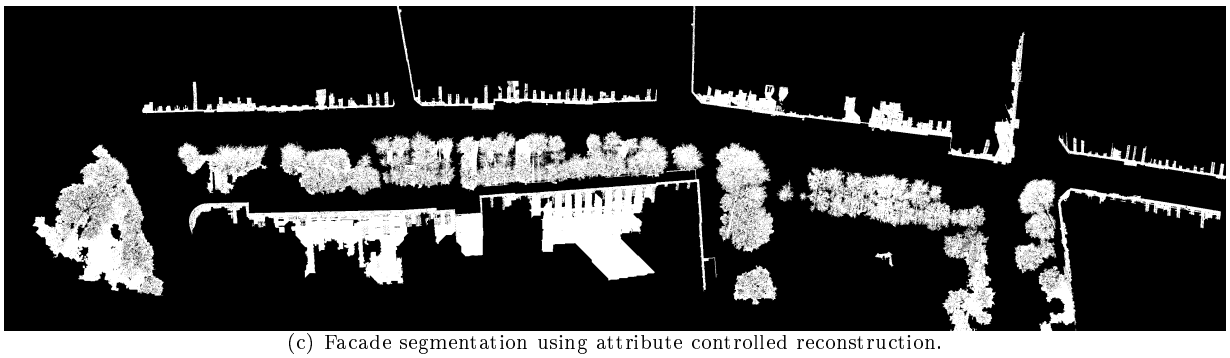
More adapted to our 3D urban data, let us to define an adaptive decomposition using slices parallel to the ground. From Definition 5.5.2, we propose the following decomposition:



(a) Elevation image.



(b) Facade markers. Several incorrect markers have been detected due to tree alignments.



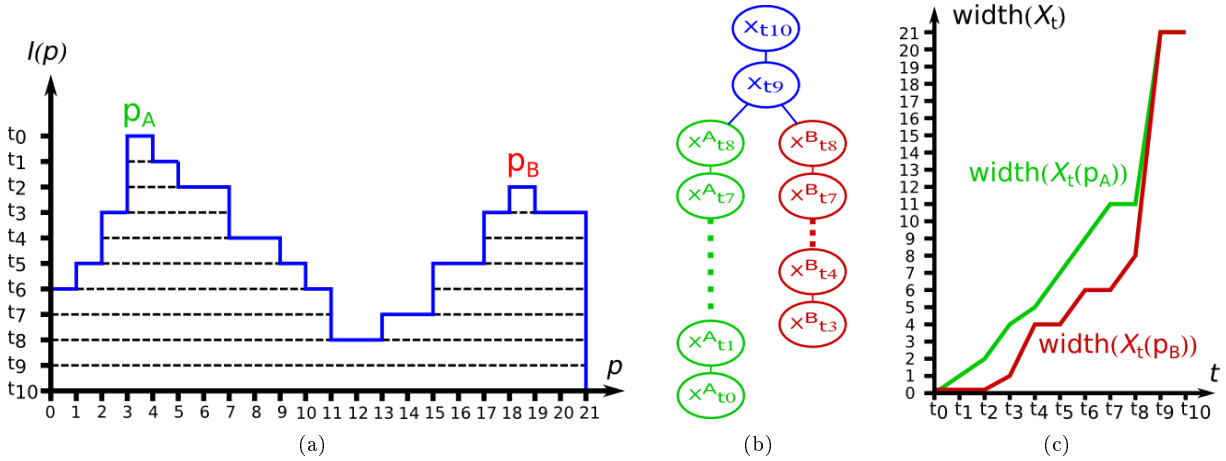
(c) Facade segmentation using attribute controlled reconstruction.

Figure 5.11: Errors in facade segmentation due to tree alignments wrongly extracted as facade markers. That is why we have proposed a more robust method without facade markers. Using such method, only the elongation and its evolution over the height decomposition of the scene are analyzed. This method is proved to produce the best results, as explained in the following section. Test site in *St. Sulpice square* in Paris, France. Stereopolis II, IGN©.

Definition 5.5.3 *Let f_{gr} be a digital gray-scale image $f_{gr} : D \to V$, with $D \subset Z^2$ the image domain and $V = [0, ..., R]$ the set of gray levels mapping the ground elevation, resulting from the ground segmentation process explained in Section 4.4. A decomposition of P using slices parallel to the ground can be obtained considering successive thresholds from the ground separated by a given height Δz:*

$$T_t^{gr}(P) = \{m_i \in D \mid f_{gr}(m_i) + t\Delta z < z_i < f_{gr}(m_i) + (t+1)\Delta z\};\ \forall t = [0, ..., R-1];\ \forall i = [0, ..., N] \quad (5.9)$$

This decomposition is equivalent to an adaptive voxelization, as shown in Figure 5.13. Dashed lines represent the slices parallel to the ground. For each slice, an occupancy grid is defined according to the elevation image pixel size $1/k$, where k is the number of pixels per unit length (For further details, see Section 3.4.2). Each voxel is labeled *full* if there is at least one 3D point inside, or *empty* otherwise. Finally, these occupancy grids are stacked in a binary 3D image. For each slice, attributes are computed on each binary CC.



(a) (b) (c)

Figure 5.12: (a) 1D threshold decomposition; (b) component tree; (c) attribute (width) profile for the two maxima (p_A and p_B). Events on this attribute profile are useful to segment objects, extract features and define adaptive structuring elements.

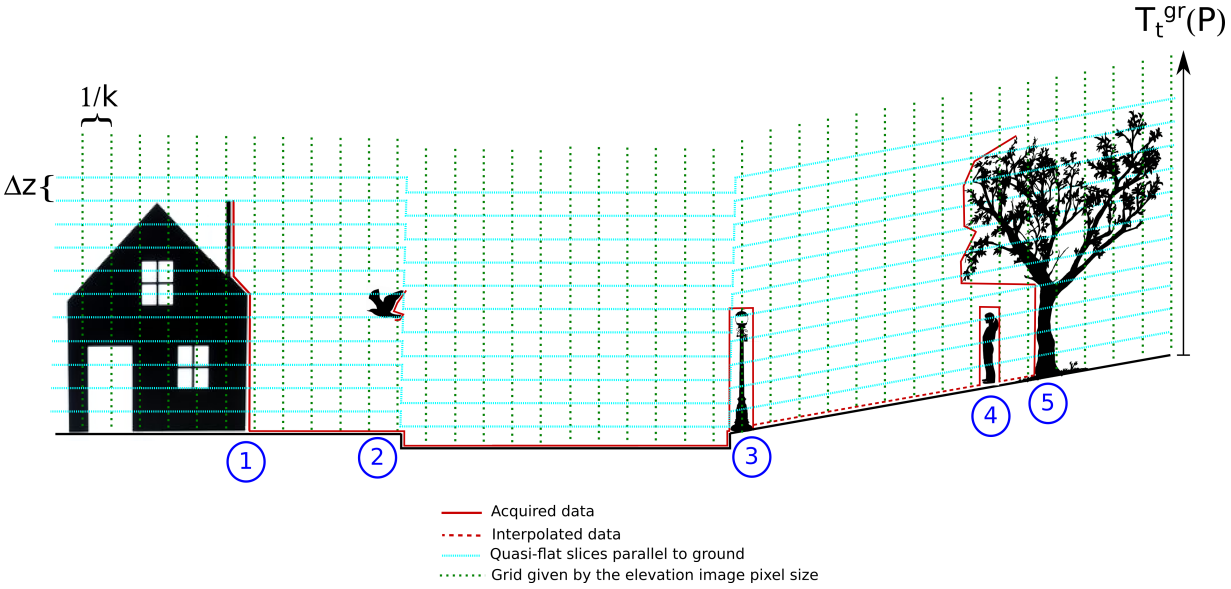


Figure 5.13: Adaptive voxelization using slices parallel to the ground. Dashed lines represent slice divisions. For each slice, an occupancy grid is defined according to the elevation image pixel size $1/k$, where k is the number of pixels per unit length. Each voxel is labeled *full* if there is at least one 3D point inside, or *empty* otherwise. Finally, these occupancy grids are stacked in a binary 3D image. For each slice, attributes are computed on each binary CC. This example contains five objects: ① facade, ② bird, ③ lamppost, ④ pedestrian, and ⑤ tree.

We propose to segment facades using the maximal elongation image computed from the attribute profile of decomposition T_t^{gr}. With this aim, we compute the geodesic elongation $E(X_t)$ for each CC on each slice parallel to the ground. Then, for each pixel m_i, we store the maximal elongation over the whole decomposition:

$$E_{max}(m_i) = max|E(X_t(m_i))|;\ \forall X_t \in T_t^{gr};\ \forall t \in [0, ..., R-1] \tag{5.10}$$

Such feature image is a partition of the space where each pixel contains information about elongation of its neighborhood. Then, it is useful in segmentation tasks where some prior shape knowledge is exploitable. This

decomposition is used to segment facades while filtering out other structures, including objects connected to it. The slice height has been set to $\Delta z = \Delta\lambda = 1.0$ m, since we are only interested in connected objects higher than 1.0 m (motorcycles, pedestrians, urban furniture, etc.). Additionally, it offers a trade off between processing time and performance, since only a few tens of slices are required to decompose an urban scenario with high buildings.

Figure 5.14 illustrates an example of facade segmentation using this approach. Figure 5.14(a) shows the elevation image. Figure 5.14(b) presents the elongation image computed from the threshold decomposition of the 3D point cloud. Figure 5.14(c) presents the segmentation result applying a simple threshold on the elongation image. We define *elongFacade* as the minimal elongation allowed for a facade. In our experiments, we have heuristically set *elongFacade*=20, which corresponds to the elongation of a rectangle whose length is 25 times longer than its width.

Figure 5.15 presents another facade segmentation result on a test site in *rue Bonaparte* in Paris, France. With respect to Figure 5.11 most of facades are correctly segmented. The only problems appear in the left part: zone 1, where the side part of a bus has been wrongly detected as facade; and zone B, where bushes and vegetation over a low wall could not be separated (Figure 5.15(d)). These objects present a high elongation, they are then segmented as facades.

5.6 City block segmentation

A city block is the smallest area that is surrounded by streets. A wide variety of sizes and shapes can be found in urban environments. In general, it depends on historic, demographic and geographic constraints. For example, many pre-industrial cities tend to have irregular city blocks, while newer cities have usually much more regular arrangements (Wikipedia, 2014).

In our application, city blocks are considered as the biggest semantic entity in the urban environment. Their segmentation is useful for individual city block analysis, e.g. occluded curbs belonging to different city blocks should not be reconnected, as explained in Chapter 4. Additionally, each city block may be processed separately and their results joined at the end of the analysis, reducing memory requirements and allowing parallelization.

Once facades have been segmented on the elevation image, we compute the influence zones (IZ) of each facade in order to define city blocks. The IZ was one of the first morphological operators applied to image segmentation. It was discovered in the 70s from the iterative application of basic operators such erosion and dilation (Matheron, 1975; Serra and Soille, 1994). The IZ of a given CC is defined by the set of pixels of a binary image that are closer to this CC than to any other CC on the image. Let us introduce its formal definition:

Definition 5.6.1 ***Influence zones (IZ).*** *Let X be a binary image and K_1, K_2,..., K_n the CC of X. The influence zone of K_i is the set of pixels of image X which is closer to K_i than any other CC of image X:*

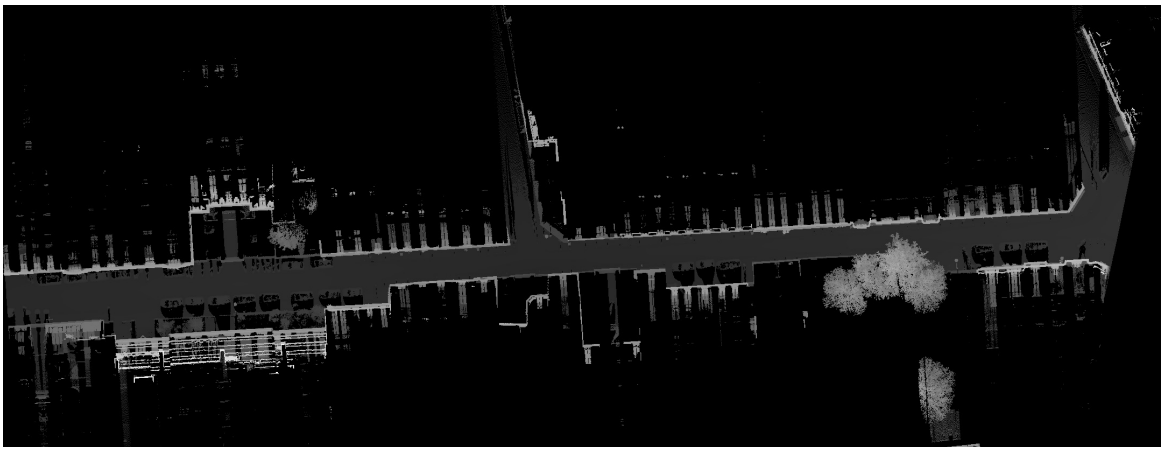
$$IZ(K_i) = \{p | \forall j \in \{1, ..., n\}, i \neq j \Rightarrow d(p, K_i) < d(p, K_j)\} \tag{5.11}$$

It is noteworthy that this is an alternative definition of a Voronoi diagram (Voronoi, 1908). In practice, the IZ of a binary image is computed using a constrained watershed on the distance function of the binary image. Figure 5.16 illustrates this city block segmentation. Figure 5.16(a) shows the elevation image. Figure 5.16(b) presents the binary image containing the facade segmentation result. Figure 5.16(c) shows a morphological closing of size *sepFacade* in order to reconnect near facades belonging to the same city block, *i.e.* *sepFacade* stands for the minimal separation between city blocks. Figure 5.16(d) shows the medial road axes useful to avoid defining city block crossing the street. This information is used if available and it can be obtained from the vehicle trajectory or from an external 2D map. Figure 5.16(e) illustrates the distance function computed from facades. The distance function is constrained to be maximum on the medial road axes and on the no-data pixels. Finally, Figure 5.16(f) presents the IZ as the result of a constrained watershed on the distance function. Each color represents a different city block.

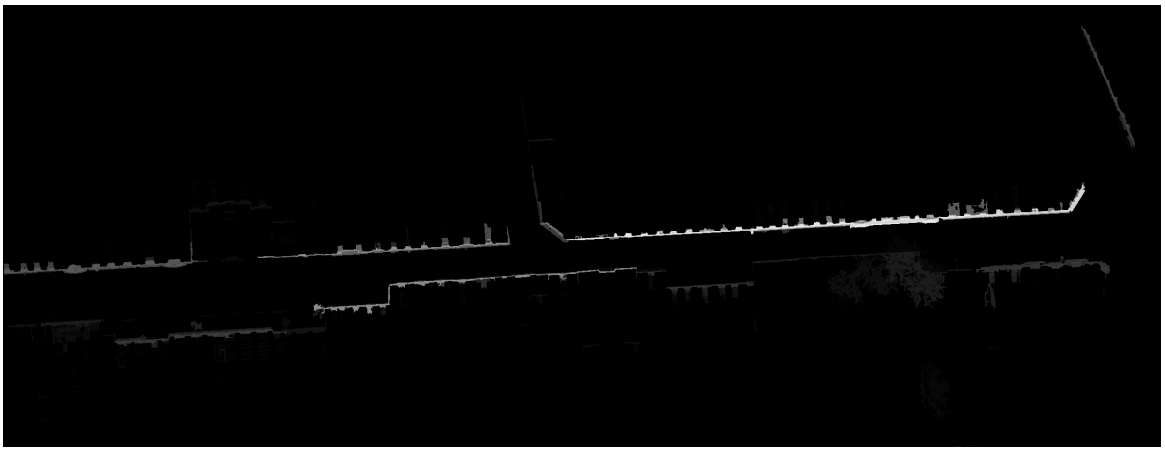
Figures 5.17 and 5.18 present two city block segmentation results reprojected onto the 3D point cloud. In those experiments, facades have been segmented using the image elongation based method (Section 5.5).

5.7 Results

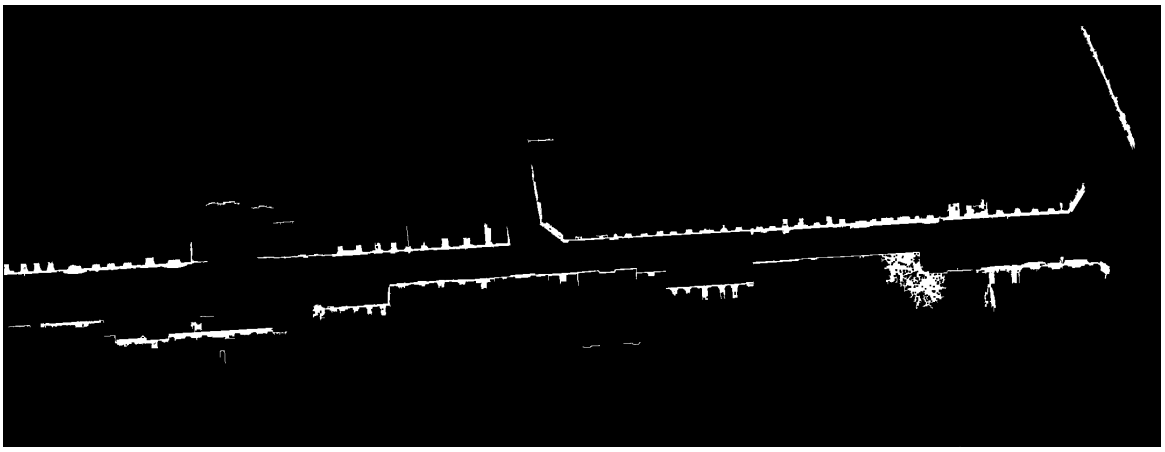
Our facade segmentation methods have been tested on TerraMobilita datasets in order to get qualitative and quantitative results. Two types of ground truth (GT) annotations are available:



(a) Elevation image.

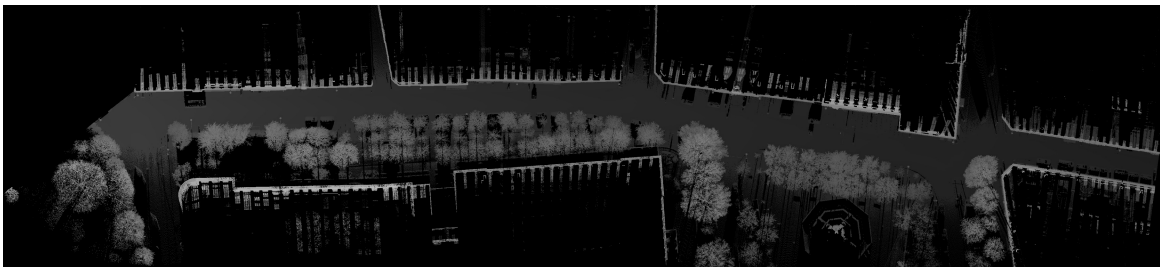


(b) Maximal elongation image computed from the 3D point cloud.

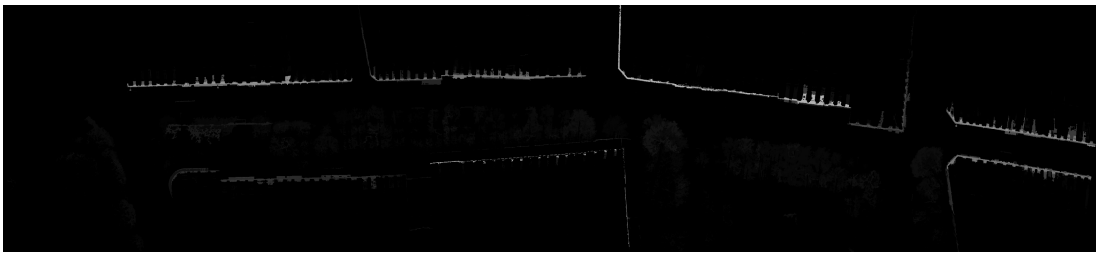


(c) Facade segmentation using the maximal elongation image.

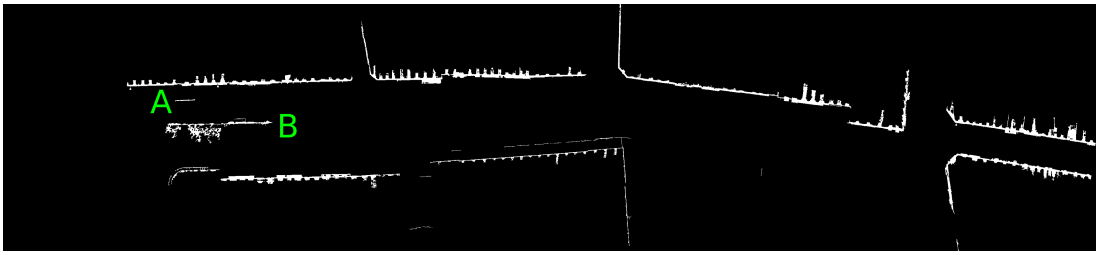
Figure 5.14: Facade segmentation using the maximal elongation image. (a) shows the elevation image. (b) presents the elongation image computed from the threshold decomposition of the 3D point cloud. (c) presents the segmentation result applying a simple threshold on the elongation image.We define *elongFacade* as the minimal elongation allowed for a facade. In our experiments, we have heuristically set *elongFacade*=20, which corresponds to the elongation of a rectangle whose length is 25 times longer than its width. Test site in *rue Cassette* in Paris, France. Stereopolis II, IGN©.



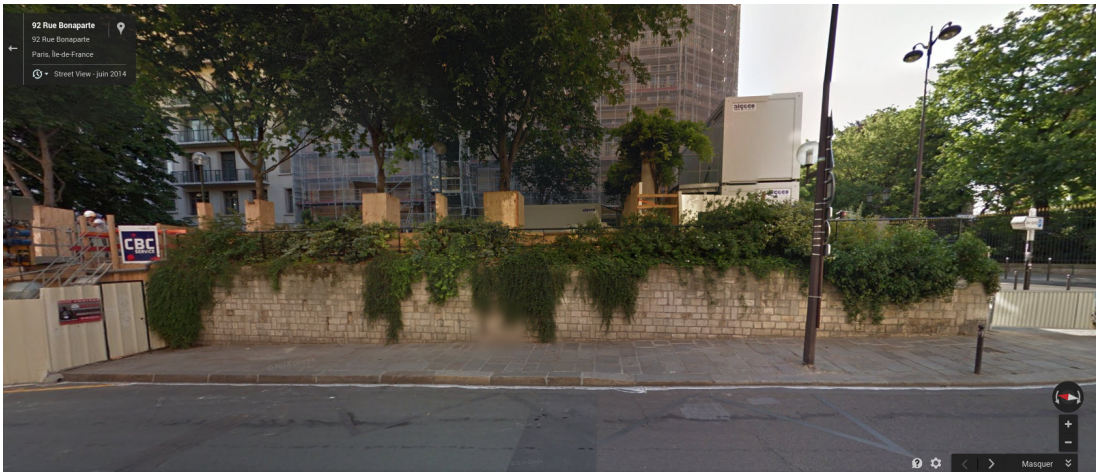
(a) Elevation image.



(b) Maximal elongation image computed from the 3D point cloud.

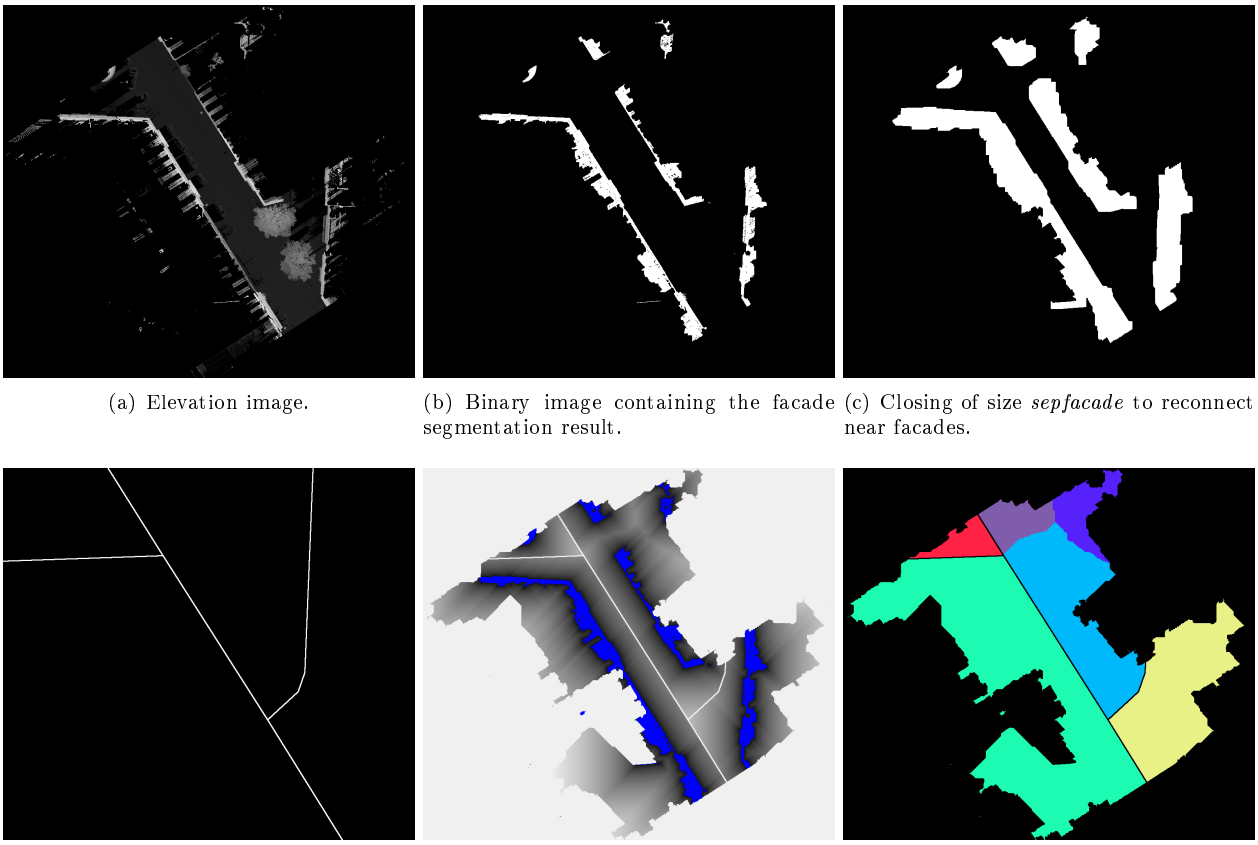


(c) Facade segmentation using the maximal elongation image. Two segmentation errors have been found in zones A and B.



(d) Facade segmentation error due to vegetation and bushes over a low wall. This is an unusual case presented in zone B in Figure 5.15(c).

Figure 5.15: Facade segmentation using the maximal elongation image. Test site in *rue Bonaparte* in Paris, France. Stereopolis II, IGN©. With respect to fig. 5.11 most of facades are correctly segmented. The main problems appear in the left part: zone 1, where the side part of a bus has been wrongly detected as facade; and zone B, where bushes and vegetation over a low wall could not be separated. These objects present a high elongation, they are then segmented as facades.



(a) Elevation image.

(b) Binary image containing the facade segmentation result.

(c) Closing of size *sepfacade* to reconnect near facades.

(d) Medial road axes (used if available).

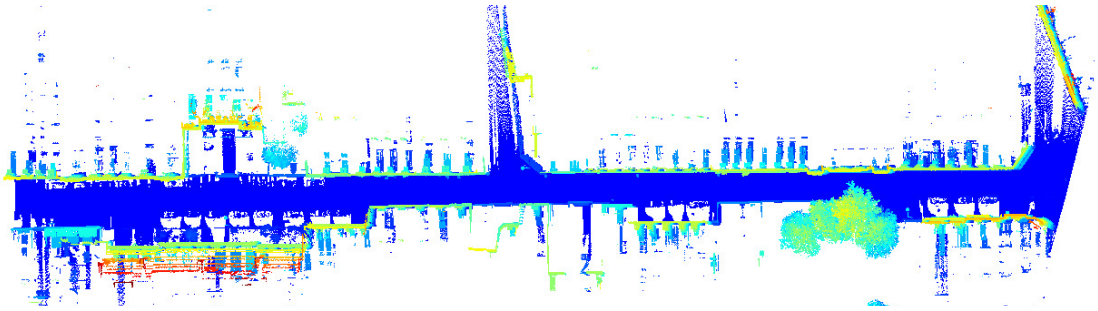
(e) Distance function from facades (blue). No-data pixels and medial road axes are constrained to have the maximal distance.

(f) Influence zones. Each color represents a different city block.

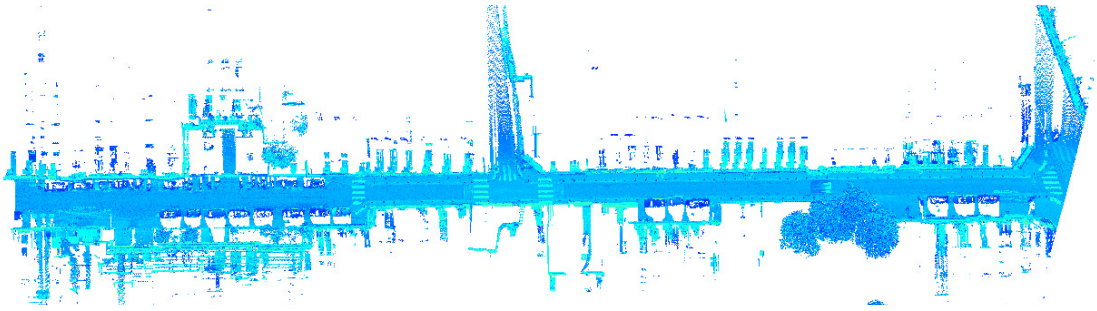
Figure 5.16: City block segmentation using the influence zones of the facade. Test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©. (a) shows the elevation image. (b) presents the binary image containing the facade segmentation result. (c) shows a morphological closing of size *sepFacade* in order to reconnect near facades belonging to the same city block. (d) shows the medial road axes useful to avoid defining city block crossing the street. (e) illustrates the distance function computed from facades. The distance function is constrained to be maximum on the medial road axes and on the no-data pixels. Finally, (f) presents the result of a constrained watershed on the distance function. Each color represents a different city block.

- 2D lines indicating the separation between sidewalks and buildings. These 2D manual annotations are usually provided by local authorities. In our case, they have been obtained from Open Data Paris (ODParis http://opendata.paris.fr/), a project from Paris city hall (*Mairie de Paris*, in French) in order to make urban data available to the community. Evaluations using 2D lines are commonly used in the state of the art when 3D annotations are not available (Vosselman and Zhou, 2009; Zhou and Vosselman, 2012; Serna and Marcotegui, 2013b). These evaluations give an idea on the segmentation method performance. However, results should be carefully interpreted since the evaluation is only carried out on the 2D space at the ground level, then performance segmenting 3D features such as inclined facades, architectural details and balconies cannot be directly evaluated.
- 3D point-wise annotations, *i.e.* a class is assigned to each 3D point. These point-wise annotations allow a global evaluation taking all facade points into account. In our opinion, this evaluation is the most appropriate, however complete 3D manual annotations are rarely available in the state of the art. In our experiments, we have used 3D annotations and evaluation methods developed in the framework of TerraMobilita/iQmulus benchmark (http://data.ign.fr/benchmarks/UrbanAnalysis/).

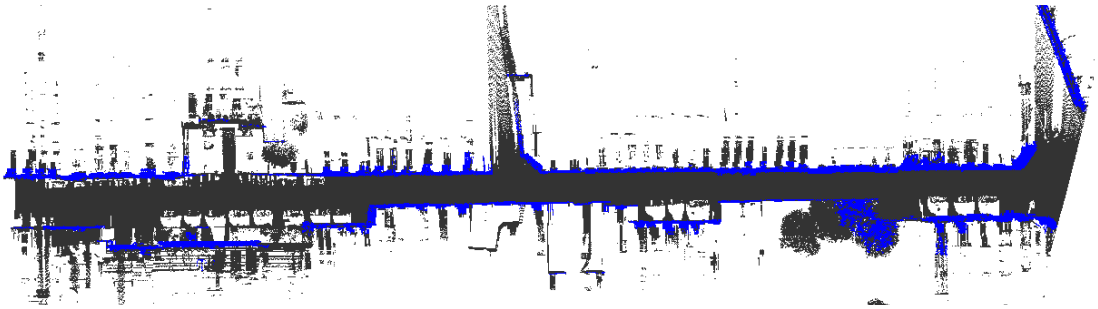
Evaluations using each type of GT annotation are presented in the two following subsections.



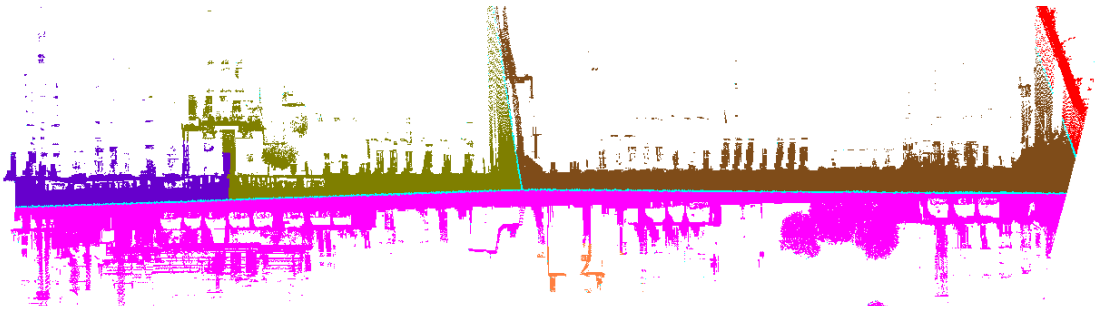
(a) Point cloud colored by the Z coordinate.



(b) Point cloud colored by the laser intensity.



(c) Facade segmentation using the elongation image (method without markers). Facade (blue) and other (gray).



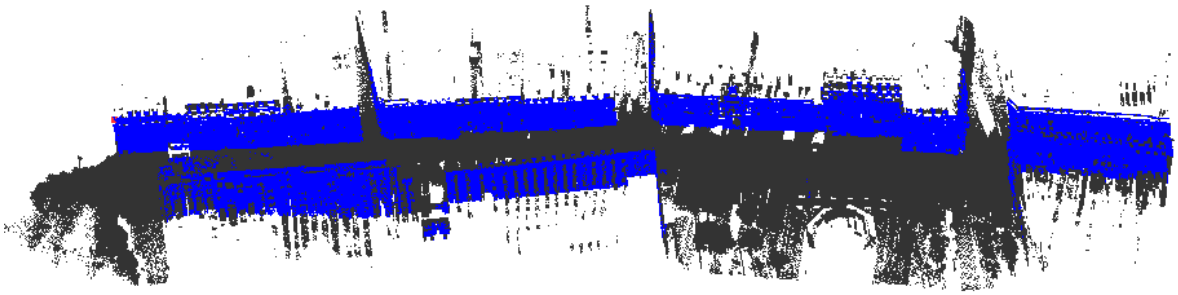
(d) City block segmentation. Each color represents a different city block.

Figure 5.17: City block segmentation using the influence zones of the facade. Reprojection onto the 3D point cloud. Test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©.

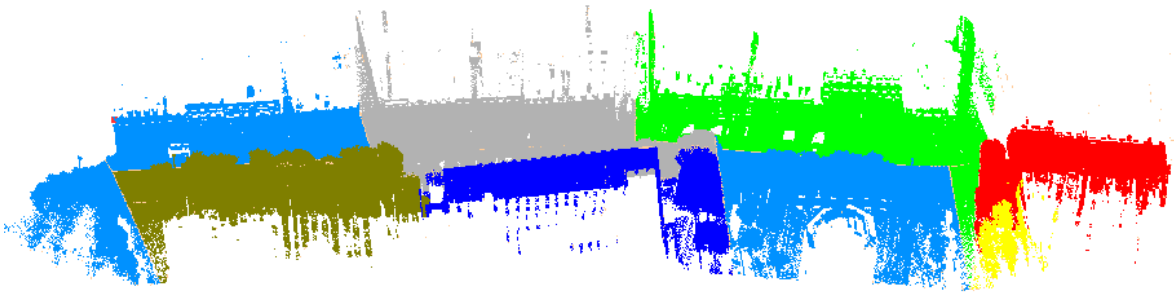
5.7.1 Results: Evaluation using Open Data Paris

We have chosen 4 MLS datasets from 6^{th} Parisian district, France. Data have been acquired by Stereopolis II, a MLS system by the National French Mapping Agency (IGN) (Paparoditis et al., 2012), previously presented in Section 2.4.1. Table 5.1 describes the datasets used in this evaluation.

GT annotations have been obtained from ODParis and correspond to 2D lines indicating the separation



(a) Facade segmentation using the elongation image (method without markers). Facade (blue) and other (gray).



(b) City block segmentation. Each color represents a different city block.

Figure 5.18: City block segmentation using the facades influence zones. Reprojection onto the 3D point cloud. Test site in *rue Bonaparte* and *St. Sulpice* square in Paris, France. Stereopolis II, IGN©.

Table 5.1: Datasets used for evaluation of our facade segmentation methods. All datasets have been acquired in January 2013 by Stereopolis II system in the 6^{th} Parisian district, France.

ID	Filename	Description
site I	TerMob2_LAMB93_0020.ply	*rue d'Assas*, approx. 90 m. Short street section with straight facades. See Figure 5.20
site II	TerMob2_LAMB93_0021.ply	*rue d'Assas*, approx. 80 m. Short street section with two intersections and two big trees. See Figure 5.21
site III	Cassette_idclass.ply	*rue Cassette*, approx. 200 m. Long street with mainly straight facades and a small tree alignment. See Figure 5.22
site IV	Z2.ply	*rue Bonaparte* and *St. Sulpice* square, approx. 400 m. Long street section with non straight facades and several tree alignments. See Figure 5.23

between sidewalks and buildings. Quantitative analyses are performed by comparison between automatic and GT lines on a 2D image. When 3D facade points are projected to a 2D plane, they are usually wider than a single line due to facade inclination, architectural details and balconies. Therefore, buffers around GT lines and segmented facades are required in order to compute the evaluation.

On the one hand, a segmented facade is labeled as true positive or false positive if it is located inside or outside a GT buffer, respectively. On the other hand, a GT facade is labeled as segmented or missed if it is located inside or outside a segmented facade buffer, respectively.

In our datasets, we consider that a buffer width of 1.0 m is appropriate to quantify true positives without overestimating false positives, as shown in Figure 5.19. This buffer-based evaluation is commonly used in other works reported in the literature (Vosselman and Zhou, 2009; Zhou and Vosselman, 2012; Serna and Marcotegui,

2013b).

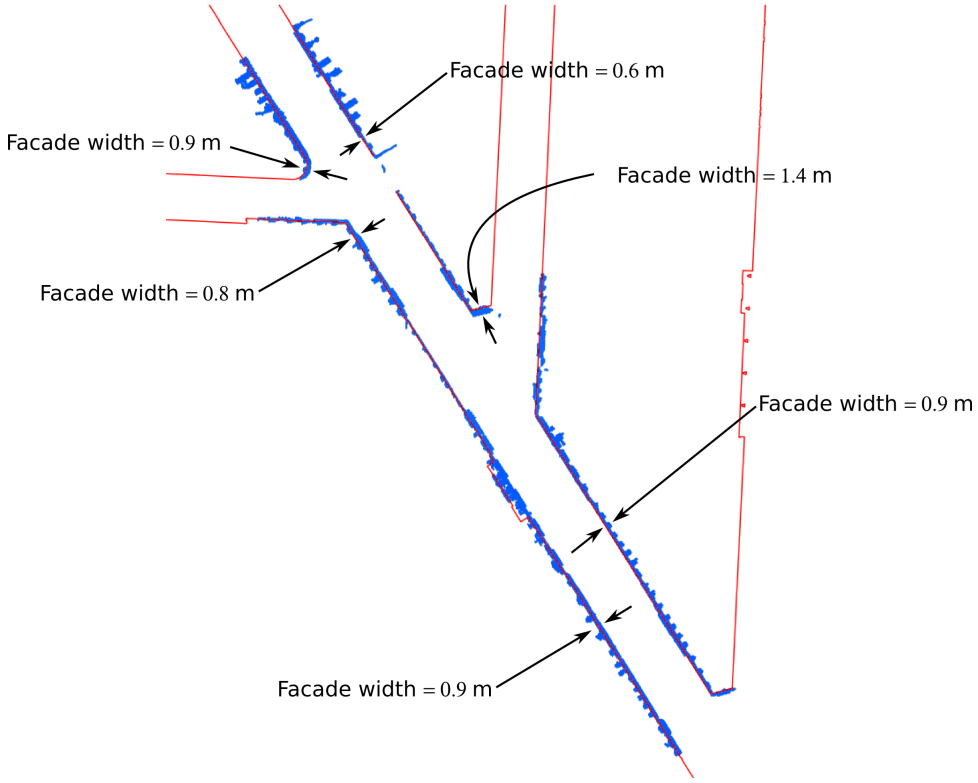


Figure 5.19: Ground truth lines and 3D facade points projected onto the 2D plane. GT annotations have been obtained from ODParis and correspond to 2D lines indicating the separation between sidewalks and buildings. GT lines (red) and segmented facades (blue). Test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©.

The classic Precision (P), Recall (R) and f_{mean} criteria are computed. Recall is defined as the number of GT pixels correctly segmented divided by the total number of GT pixels; Precision is defined as the number of true positive pixels divided by total number of segmented pixels (true positives + false positives); and $f_{mean} = 2PR/(P+R)$.

Table 5.2 presents a quantitative comparison between our facade segmentation methods. As aforementioned, these results should be carefully interpreted since the evaluation is only carried out on the 2D space at the ground level. Then, the performance segmenting inclined facades, facades with architectural details and balconies cannot be directly quantified.

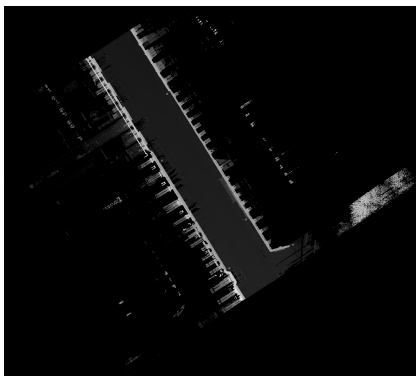
Figures 5.20 to 5.23 present the elevation images, the GT annotations and the segmentation results on the four test sites. It is noteworthy that several ground truth annotations are missing. For example, in site III (Figure 5.22(b)) several facades in the street left side are missing, while in site IV (Figure 5.23(b)) facades behind trees in the street right side have not been annotated. Therefore, several correct segmentations have been incorrectly labeled as false positives. As a result, the performance of our methods is sub-estimated in these two test sites. Let us analyze each method individually.

Method 1, based on reconstruction by dilation from markers, presents the highest Recall retrieving 100% of facades in the four test sites. However, this method presents also the highest number of false positives (Precision ranges between 13.6% and 45.1% for all test sites). As aforementioned, this method is based on iterative geodesic dilation, then any object touching the facade is segmented as part of it. This method is the fastest one and its use may be justified in an application with strict time constraints or if only a rough segmentation is required. For example, if we are only interested in defining the public space boundary (*e.g.* for a urban mobility application), all objects touching or behind the facade are not required to be segmented.

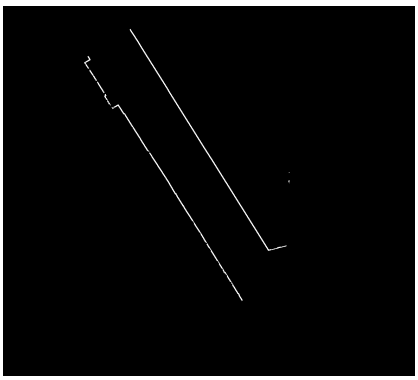
As aforementioned, the main problem of this method is that connected objects, such as motorcycles parked next to the facade or leaning pedestrians, are reconstructed in the facade mask. In order to solve this problem, we have proposed Method 2, based on attribute controlled reconstruction from markers. Since connected objects

Table 5.2: Quantitative comparison between our facade segmentation methods on 4 test sites.

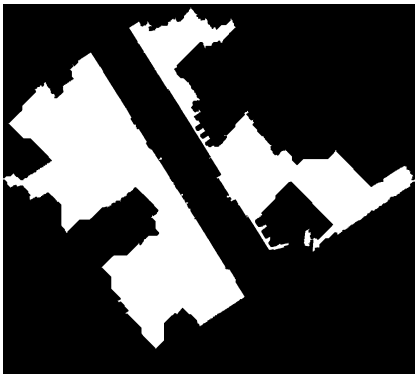
ID	Method 1: reconstruction by dilation (Section 5.4)	Method 2: attribute controlled reconstruction (Section 5.4)	Method 3: maximal elongation image (Section 5.5)
site I	Precision = 45.1% Recall = 100.0% f_{mean} = 62.2% time = 2.2 s	Precision = 73.2% Recall = 99.5% f_{mean} = 84.4% time = 10.7 s	Precision = 80.9% Recall = 99.3% f_{mean} = 89.2% time = 20.0 s
site II	Precision = 47.7% Recall = 100.0% f_{mean} = 64.6% time = 2.1 s	Precision = 87.8% Recall = 97.9% f_{mean} = 92.6% time = 7.8 s	Precision = 92.3% Recall = 93.2% f_{mean} = 92.8% time = 16.5 s
site III	Precision = 27.5% Recall = 99.3% f_{mean} = 43.0% time = 3.5 s	Precision = 44.8% Recall = 98.6% f_{mean} = 61.6% time = 22.5 s	Precision = 67.9 % Recall = 98.2% f_{mean} = 80.3 % time = 44.6 s
site IV	Precision = 13.6% Recall = 100.0% f_{mean} = 23.8% time = 6.8 s	Precision = 13.9% Recall = 99.8% f_{mean} = 27.4% time = 68.6 s	Precision = 64.7% Recall = 88.9% f_{mean} = 74.9% time = 108.4 s



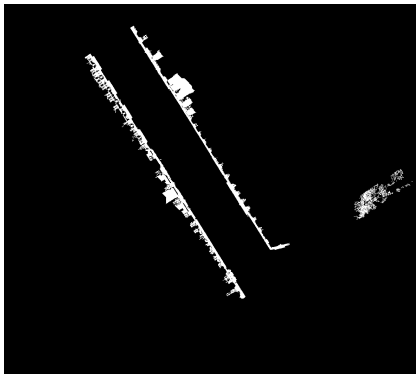
(a) Elevation image.



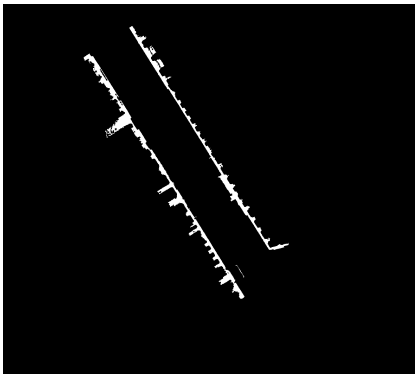
(b) Ground truth. Facade lines taken from ODParis.



(c) Facade segmentation using reconstruction by dilation.

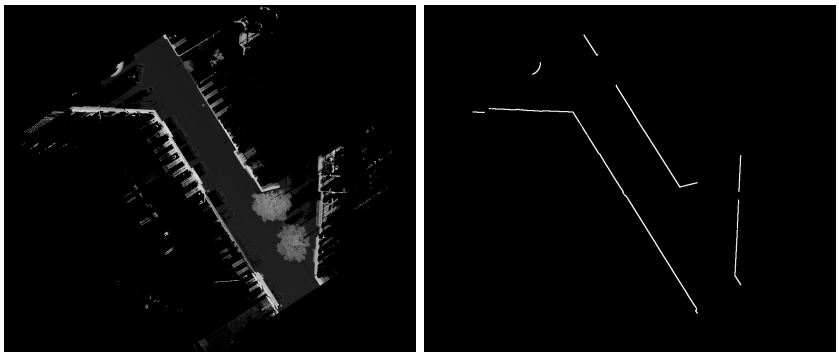


(d) Facade segmentation using attribute controlled reconstruction.



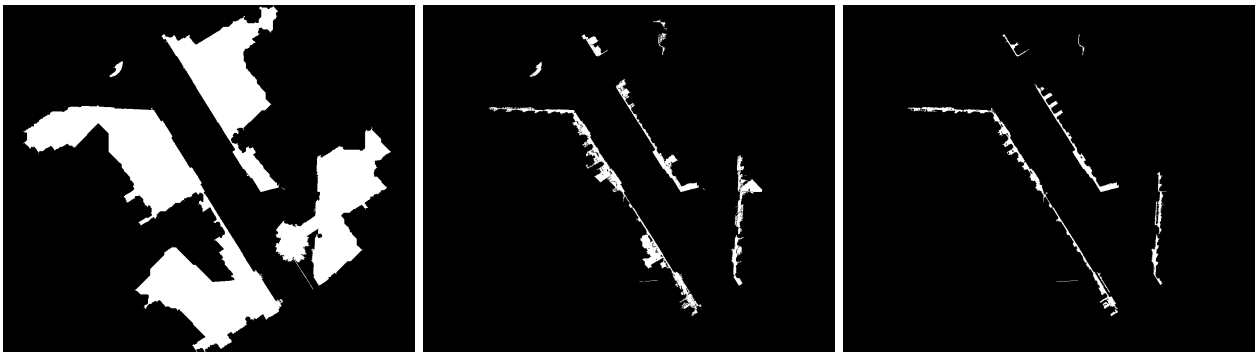
(e) Facade segmentation using the maximal elongation image.

Figure 5.20: Facade segmentation results for site I (TerMob2_LAMB93_0020.ply). (a) presents the elevation image, (b) the GT annotations and (c,d,e) the segmentation results using our three proposed methods.



(a) Elevation image.

(b) Ground truth. Facade lines taken from ODParis.



(c) Facade segmentation using reconstruction by dilation.

(d) Facade segmentation using attribute controlled reconstruction.

(e) Facade segmentation using the maximal elongation image.

Figure 5.21: Facade segmentation results for site II (TerMob2_LAMB93_0021.ply). (a) presents the elevation image, (b) the GT annotations and (c,d,e) the segmentation results using our three proposed methods.

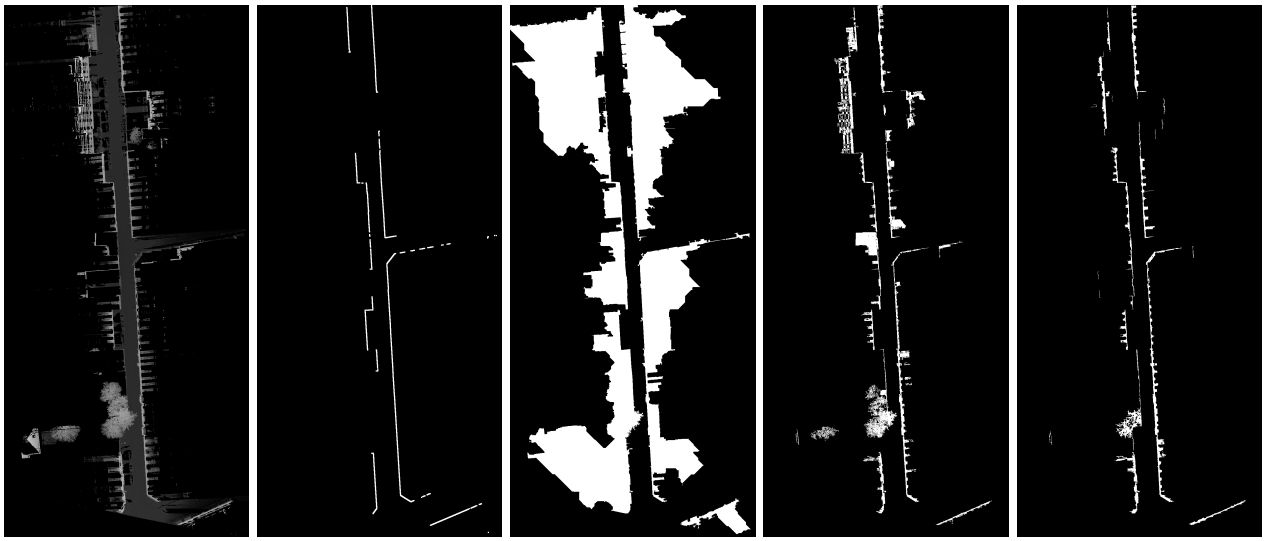
usually reduce the global facade elongation, this method offers better results than the first one: Recall is higher than 97% for all test sites while Precision increases up to 73.2% and 87.8% in sites I and II, respectively. This method presents the best trade-off between performance and processing time.

In general, methods based on facade markers are strongly influenced by the markers selection method. The main drawback is that bad located markers may produce errors reconstructing non-facade objects. In particular in site IV, marker-based methods fail segmenting the tree alignments in the street right side (Precision is 13.6% and 13.9% for methods 1 and 2, respectively). In order to avoid markers, we propose Method 3, a more robust segmentation method based on the maximal elongation image. This method is proved to produce the best results for all test sites: f_{mean} equal to 80.9% and 92.8% for sites I and II. In spite of missing GT annotations, f_{mean} is equal to 80.3% and 74.9% for sites III and IV, proving the performance of this method even in the presence of trees. The main drawback is that its implementation is slow, then it is not suitable for real-time applications. However, it remains possible for large scale applications, where time constraints are less strict. Note that processing time is only a few tens of seconds for an acquisition of several hundreds of meters, using a non-optimized implementation.

5.7.2 Results: TerraMobilita/iQmulus database

TerraMobilita/iQmulus database (Brédif et al., 2014) has been developed aiming at benchmarking semantic analysis methods working on 3D dense urban data. This database has been created in the framework of TerraMobilita project. It consists in 11 annotated 3D point clouds acquired by Stereopolis II system (IGN©) in the 6^{th} Parisian district in January 2013. Annotation has been carried out in a manually assisted way by MATIS laboratory at IGN. Further details on this database can be found in Section 2.6.2.

For external reasons (annotation problems and benchmark deadlines), our evaluation only consists in one of



(a) Elevation image. (b) Ground truth. Facade lines taken from ODParis. (c) Facade segmentation using reconstruction by dilation. (d) Facade segmentation using attribute controlled reconstruction. (e) Facade segmentation using the maximal elongation image.

Figure 5.22: Facade segmentation results for site III (Cassette_idclass.ply). (a) presents the elevation image, (b) the GT annotations and (c,d,e) the segmentation results using our three proposed methods.

the ten zones. For this experiment, file "Cassette_idclass.ply" has been used[1]. It contains 12 million points from a street section approximately 200 m long in *rue Cassette* in Paris, France. Manual annotations and point-wise evaluations have been independently carried out by the National French Mapping Agency (IGN). Further details on this annotation and this evaluation will be presented later in Section 6.7. Results of this benchmark have been presented on July 8th, 2014 in Cardiff (UK), in conjunction with SGPâĂŹ14 (Vallet et al., 2014).

Figure 5.24 presents the facade segmentation result projected onto the 3D point cloud. In this experiment, only our method based on the maximal elongation image has been applied. As a general remark, errors of our segmentation method are due to an incomplete detected facade (zone A) and a tree alignment connected to a low wall (zone B).

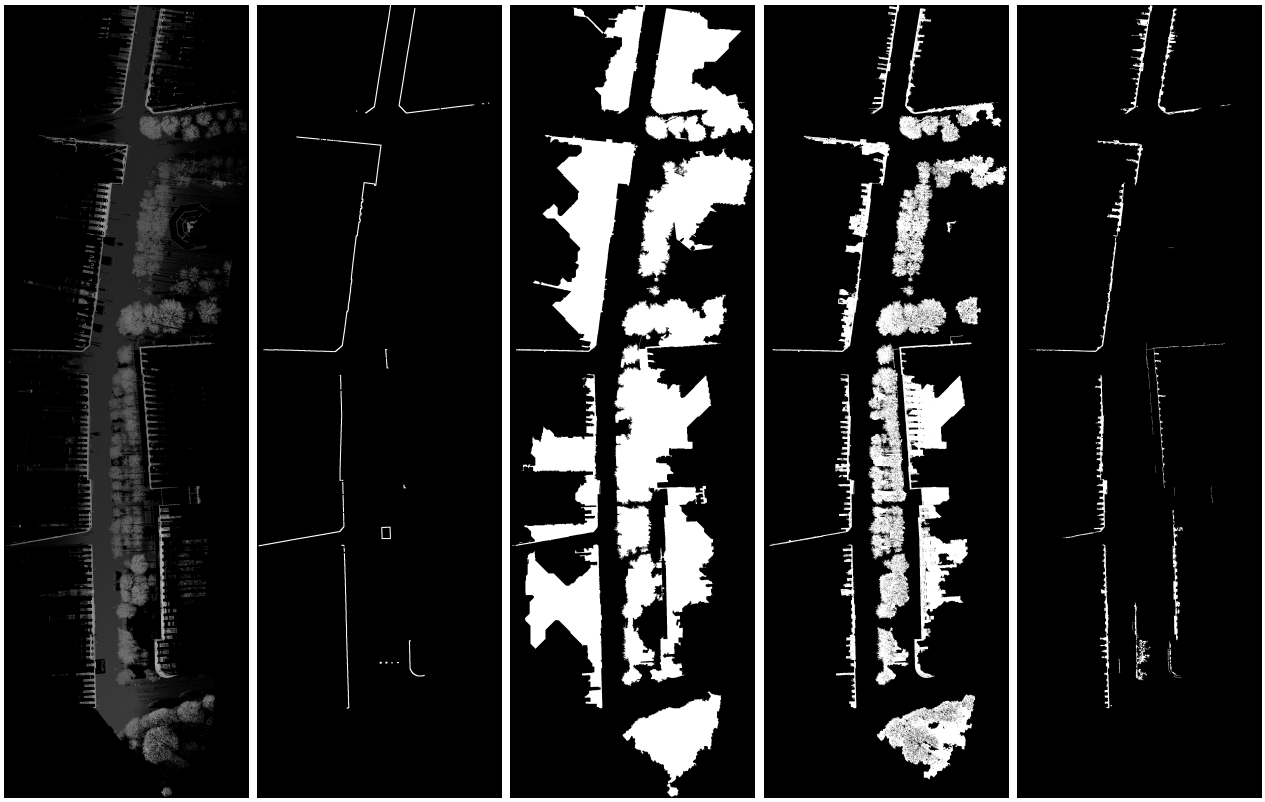
As aforementioned, our results are evaluated point-by-point using the TerraMobilita/iQmulus evaluation protocol (Brédif et al., 2014). First, we classify the 3D point cloud in 3 main categories: *surface* (containing facades and ground), *object* and *other*. Moreover, the *unclassified* category has been defined for non-annotated points in the GT, which are ambiguous points difficult to annotate. They correspond to 18.31 % of total number of points in the dataset. For example, consider the tree and the wall in zone C in Figure 5.24. These points have been manually marked as *unclassified*, then they have not been taken into account in the evaluation.

Table 5.3 presents the confusion matrix and our classification results for these 3 categories. This classification is useful to evaluate the ability of our method segmenting surfaces (facade and ground) while separating objects connected to them.

Using our method, the f_{mean} for the *surface* class is equal to 96.03% while *objects* are correctly separated from them with f_{mean} equal to 84.59 %. In this experiment, we are mainly interested in separating facades and ground from other structures such as connected objects. Note that the *surface* class includes facades and ground, which represent the biggest categories in the scene with 75.82 % of total 3D points, while the *object* class represents 5.7 % of total 3D points. The overall accuracy of our method considering these categories is 92.65 %.

Table 5.4 presents our segmentation results for the *surface* class. Note that our method correctly separates facades and ground giving f_{mean} equal to 97.25 % and 98.72 %, respectively. Figure 5.25 shows that small errors are due to the facade-ground junction, where some points may be wrongly assigned. The overall accuracy

[1] The manual annotated 3D point cloud is available at:
http://data.ign.fr/benchmarks/UrbanAnalysis/download/Cassette_idclass.zip
The 3D point cloud processed by our method is available at:
https://partage.mines-telecom.fr/public.php?service=files&t=294aed38d48c8ddd03a528069f1b2e51



(a) Elevation image. (b) Ground truth. Facade lines taken from ODParis. (c) Facade segmentation using reconstruction by dilation. (d) Facade segmentation using attribute controlled reconstruction. (e) Facade segmentation using the maximal elongation image.

Figure 5.23: Facade segmentation results for site IV (Z2.ply). (a) presents the elevation image, (b) the GT annotations and (c,d,e) the segmentation results using our three proposed methods.

Table 5.3: Classification results for 3 general categories on TerraMobilita/iQmulus database. GT: ground truth, AR: automatic result. In the confusion matrix, results are presented as percentages with respect to the total number of points in the 3D point cloud (12 million points).

GT/AR	**unclassified**	**other**	**surface**	**object**	**Sum**	Recall	Precision	f_{mean}
unclassified	-	-	-	-	18.31 %	-	-	-
other	0.00 %	0.00 %	0.13 %	0.04 %	0.17 %	0.59 %	0.05 %	0.08 %
surface	1.90 %	2.19 %	70.81 %	0.91 %	75.82 %	93.40 %	98.82 %	96.03 %
object	0.09 %	0.02 %	0.72 %	4.88 %	5.70 %	85.49 %	83.72 %	84.59 %
Sum	1.99 %	2.21 %	71.66 %	5.82 %	81.69 %	**Overall accuracy: 92.65 %**		

in this case is 98.26 %. These results prove the performance of our method.

Table 5.4: Evaluation taking into account only the surface class (facades and ground) on TerraMobilita/iQmulus database. GT: ground truth, AR: Automatic result. In the confusion matrix, results are presented as percentages with respect to the total number of points in the 3D point cloud (12 million points).

GT/AR	**ground**	**facade**	**Sum**	Recall	Precision	f_{mean}
ground	30.77 %	0.01 %	30.78 %	99.96 %	94.69 %	97.25 %
facade	1.73 %	67.49 %	69.22 %	97.51 %	99.98 %	98.72 %
Sum	32.50 %	67.50 %	100.0 %	**Overall accuracy: 98.26 %**		

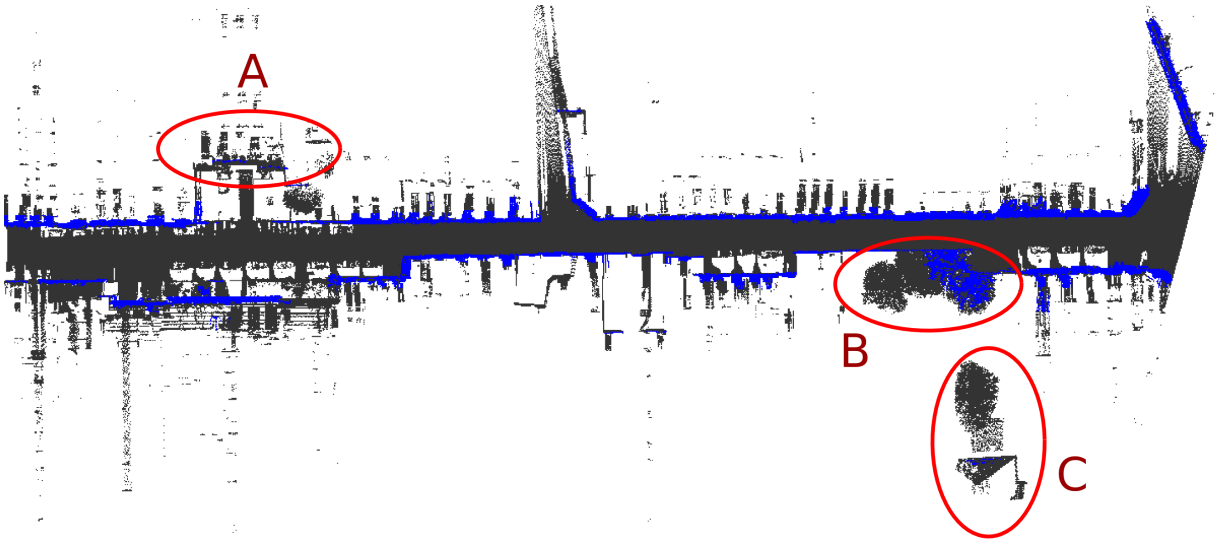
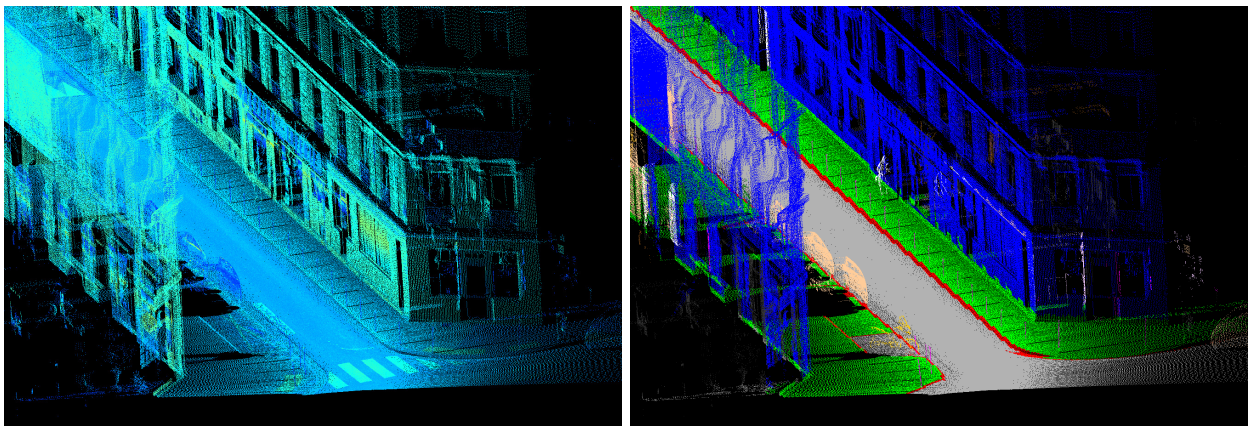


Figure 5.24: Facade segmentation result using the maximal elongation image on "Cassette_idclass.ply' file. Ground (gray), facades (blue). Input file taken from TerraMobilita/iQmulus database. Stereopolis II, IGN©. Errors are due to an incomplete detected facade (zone A) and a tree alignment connected to a low wall (zone B). Zone C corresponds to non-annotated points.



(a) Point cloud colored by the laser intensity.

(b) Segmentation result projected onto the 3D point cloud. Facade (blue), sidewalk (green), road (gray), curbs (red), pole-like objects (magenta), other (pink).

Figure 5.25: Minor errors in the facade–ground junction, where several 3D points have been wrongly assigned (between blue and green colors). Acquisition by IGN©France.

5.8 Conclusions

In this chapter, we have proposed automatic and robust approaches to segment facades from 3D point clouds. Processing is carried out using elevation images and 3D decomposition, and the final result can be reprojected onto the 3D point cloud for visualization or evaluation purposes.

Our methods are based on geometrical and geodesic constraints. Most parameters have been set heuristically and are related to urban and architectural constraints. Thus, they are intuitive to tune. The performance of our methods have been proved in our experiments on TerraMobilita databases using 2D and 3D ground truth annotations. Three approaches have been proposed: reconstruction by dilation from markers, attribute controlled reconstruction from markers and based on the maximal elongation image (without markers). Method based on reconstruction by dilation from markers is the fastest since it is based on simple thresholds and use

reconstructions constrained by the ground in order to get the entire facade. The main problem is that objects connected to the facade are reconstructed as well. In order to solve this problem, we have proposed an attribute controlled reconstruction using the geodesic elongation. Since connected objects usually appear at low heights and reduce the global facade elongation, this method offers better results than the first one. In our experiments, we have used geometric and geodesic constraints in order to extract facade markers. In the case of low facades or when the laser sensor is oriented to the ground, additional markers based on the laser rotation have been added.

In general, methods based on facade markers are strongly influenced by the markers extraction method. The main drawback is that bad located markers produce errors since they may reconstruct non-facade objects. For this reason, we have proposed a more robust method avoiding the use of facade markers. In such method, only the elongation and its evolution over the height decomposition of the scene are analyzed. This method is based on the maximal elongation image computed from 3D decomposition. It has been proved to produce the best results. However, its implementation is slower, then it is not suitable for real-time applications. Nevertheless, it remains suitable for large scale applications, where time constraints are less strict. Note that processing time is only a few tens of seconds for an acquisitions of several hundreds of meters, using a non-optimized implementation.

The selection of the best facade segmentation method remains application dependent. It should be a trade off between quality results and computational cost. In the case of a large-scale application, where time constraints are less strict, the most accurate method should be preferred.

Our approach is a research prototype, mainly based on Morph-M library (CMM, 2013), the image processing library of our laboratory. This library allows fast prototyping but it is not intended to be a fast library. Currently, the optimization of our base operators (erosion, dilation, opening, reconstruction, watershed, and so on) is under development at CMM, to bring optimized operators for real time and/or big image developments. Software (hierarchical queues, structuring elements decomposition, among others) and hardware (SIMD-Single Instruction Multiple Data and parallelization) optimizations are being integrated in SMIL library (Faessel and Bilodeau, 2013) and will be integrated in our future developments.

6 Semantic analysis of 3D urban objects

6.1 Résumé

Dans ce chapitre, nous présenterons notre analyse sémantique d'objets urbains 3D. Dans un premier temps, nous présenterons une révision de l'état de l'art. Dans un deuxième temps, nous exposerons nos algorithmes de détection, de segmentation et de classification d'objets 3D basés sur la morphologie mathématique et l'apprentissage artificiel. Dans un troisième temps, avec la coopération de l'Institut Géographique National (IGN), nous proposerons un protocole d'évaluation 3D de nos résultats. Finalement, nous reporterons des résultats quantitatifs sur des bases de données créées dans le cadre du projet TerraMobilita ainsi que sur d'autres disponibles dans la littérature.

6.2 Introduction

Digital 3D city models containing semantic object information (*e.g.* cars, pedestrians, traffic lights, trees, lampposts, etc.) are useful for many applications such as urban planning, emergency response simulation, cultural heritage documentation, virtual tourism, route planning, studies of accessibility for disabled people, parking statistics, among others.

We focus on a semantic analysis including detection, segmentation and classification of urban objects from 3D laser scanning data. In the scientific community several definitions can be found for these concepts. For the sake of clarity, let us define them in the way they should be understood in the present chapter:

Detection: An object is considered to be correctly detected if it is included in the list of object hypotheses, *i.e.* it has not been suppressed by any filtering method and it has not been included as part of the ground mask. Note that an object hypothesis may contain several connected objects or even contain only a part of an object. In the detection step, we are only interested in keeping all possible objects. This is important because in most works reported in the literature, non-detected objects cannot be recovered in subsequent algorithm steps.

Segmentation: An object is considered to be correctly segmented if it is correctly isolated as a single object, *i.e.* connected objects are correctly separated, there is no under-segmentation, and each individual object is entirely inside of one and only one connected component (CC), there is no over-segmentation. This is important because many algorithms based on clustering and connected filters can wrongly gather objects touching each other, *e.g.* motorcycles parked next to the facade, pedestrians walking together, cars closely parked to others, etc. In the segmentation step, a unique identifier (*id*) is assigned to each individual object.

Classification: In the classification step, a category (called also *class*) is assigned to each segmented object. Each *class* represents an urban semantic entity. Depending on the application, several classes can be defined: facade, ground, curbstone, pedestrian, car, lamppost, etc.

We propose an automatic semantic analysis of 3D urban objects based on elevation images, mathematical morphology and supervised learning. Our general work–flow is shown in Figure 6.1. The input is a 3D point cloud. The first three steps are presented in other chapters of this thesis: i) the 3D point cloud is projected to elevation images (presented in Section 3.4); ii) a digital terrain model (DTM) is automatically created as a result of our ground segmentation method (explained in Section 4.4); iii) facades are automatically segmented as the highest vertical structures in the elevation image (explained in Chapter 5). Then, the following three steps consist in methods for automatic detection, segmentation and classification of urban objects, and constitute the contribution of the present chapter: iv) object hypotheses are generated as discontinuities on the ground, then small and isolated regions are eliminated; v) connected objects are segmented in order to assign a unique identifier (*id*) to each individual object; vi) several geometrical and contextual features are computed for each object and classification is carried out using a Support Vector machine (SVM) approach.

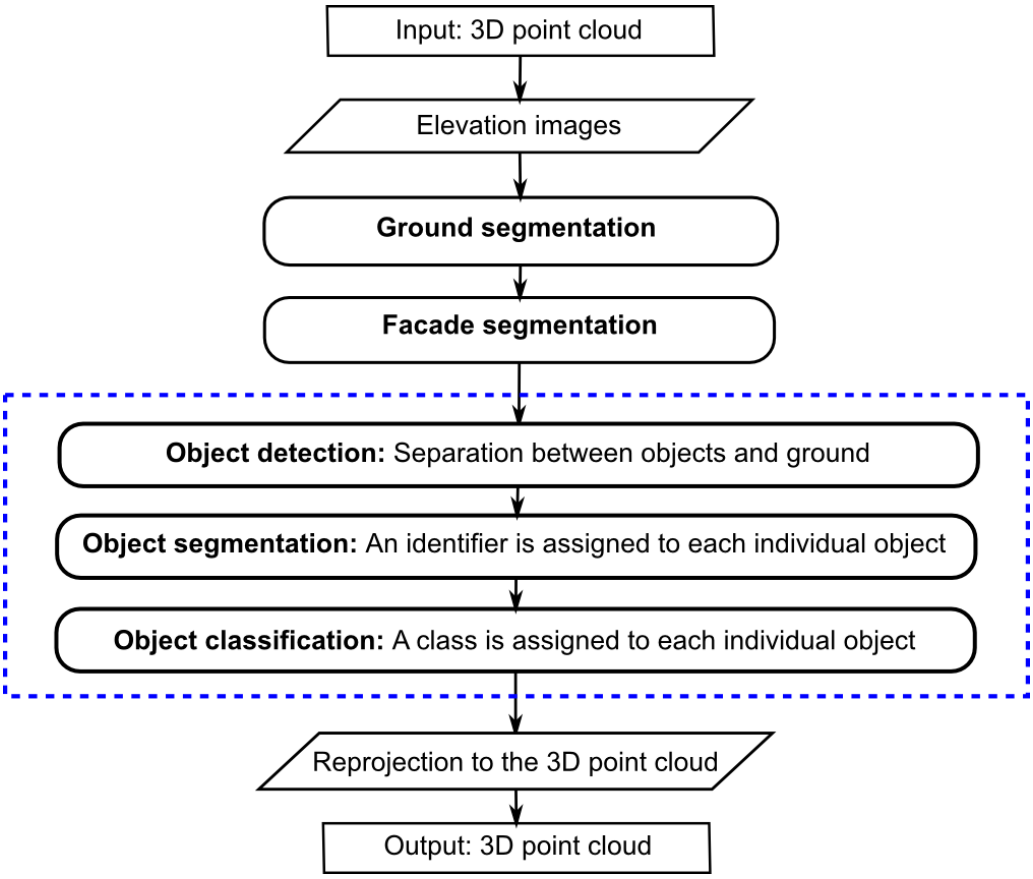
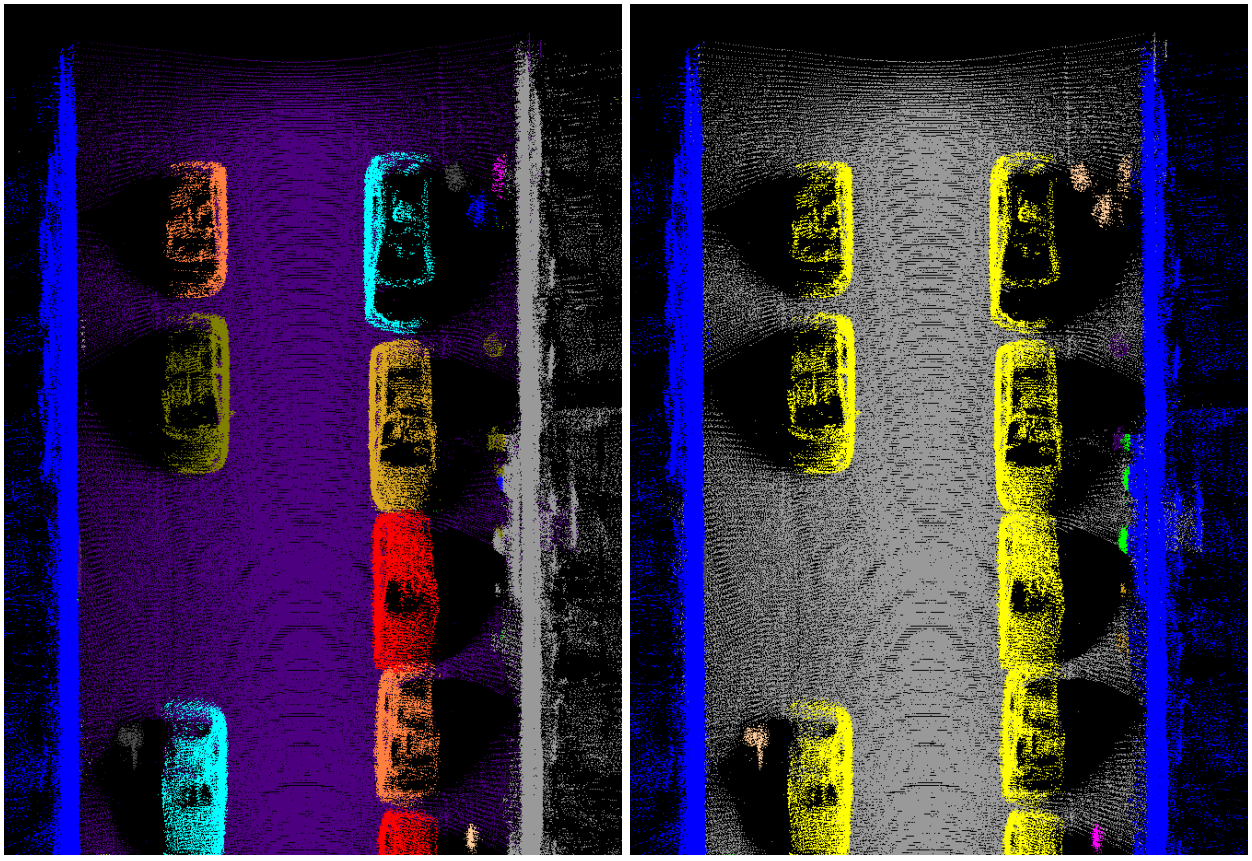


Figure 6.1: Work–flow of our proposed semantic analysis from 3D laser scanning data. Dotted blue line indicates the contributions of the present chapter on the detection, segmentation and classification of urban objects. The input is a 3D point cloud. The first three steps are presented in other chapters of this thesis: i) the 3D point cloud is projected to elevation images (presented in Section 3.4); ii) a digital terrain model (DTM) is automatically created as a result of our ground segmentation method (explained in Section 4.4); iii) facades are automatically segmented as the highest vertical structures in the elevation image (explained in Chapter 5). Then, the following three steps consist in methods for automatic detection, segmentation and classification of urban objects, and constitute the contribution of the present chapter: iv) object hypotheses are generated as discontinuities on the ground, then small and isolated regions are eliminated; v) connected objects are segmented in order to assign a unique identifier (*id*) to each individual object; vi) several geometrical and contextual features are computed for each object and classification is carried out using a Support Vector machine (SVM) approach.

As a result of our semantic analysis, two images containing *ids* and *classes* of each individual object are created. If the result have to be displayed in 3D, *id* and *class* images can be reprojected onto the 3D point cloud. For this purpose, all 3D points projected on a given pixel take the *id* and the *class* from that pixel. Having these pieces of information in two different images is useful in the case of connected objects belonging to the same class. For example, consider the alignment of parked cars shown in Figure 6.2. According to the *class* image (Figure 6.2(b)), it could be a long car parked in the right street side. However, the *id* image (Figure 6.2(a)) allows to count the number of parked cars together.

Detailed descriptions are presented in following subsections. Several contributions of this chapter have already been published in Serna and Marcotegui (2014).

This chapter is organized as follows. Section 6.3 reviews related works in the state of the art and discusses their differences with respect to our proposed methods. Sections 6.4 to 6.6 respectively introduces our detection, segmentation and classification methods based on mathematical morphology and supervised learning. Section 6.7 describes the evaluation protocol developed in the framework of TerraMobilita/iQmulus bench-



(a) Object identifier (*id*). (b) Object *class*.

Figure 6.2: *ids* and *classes* for an alignment of cars in Paris-rue-Madame dataset. For object *id*: each color represents a different object (some colors may look similar when displaying). For object *class*: facades (blue), ground (gray), cars (yellow), pedestrians (pink), urban furniture (cyan), traffic signs (red). Data acquired by L3D2, MINES ParisTech©France.

mark. Section 6.8 presents quantitative and comparative results on several state of the art databases. Finally, Section 6.9 concludes this chapter.

6.3 Related work

Even though 3D acquisition systems have a high maturity level, 3D automatic analysis of urban areas is still an active research area. In the last years, several automatic solutions have been developed with different aims.

Table 6.1 summarizes representative papers related to our work. Semantic analysis methods and performances reported on each paper are summed up in the table. Performance ranges from 58% to 95% but results are not comparable because they use different databases, consider different object classes, have different aims, use different data structures and process data in different ways. This table only offers an idea on each method performance. As a general remark, several authors use elevation images, clustering methods and supervised classifiers. Further details are given below.

Several methods project 3D information onto a 2D grid in order to reduce the problem complexity and to speed up the computational processing. As each pixel of the projected grid contains elevation information, it is called elevation image or digital elevation model. This kind of 2.5D image has a long tradition in the scientific community (Hoover et al., 1996). Besides, it is of great interest nowadays due to technological developments in remote sensing equipment such as Riegl, Velodyne and Kinect sensors since 3D points can be projected to 2D grids for visualization and processing purposes.

Gorte (2007) presents a method to segment planes on Terrestrial Laser Scanning (TLS) data using range

Table 6.1: Comparison of detection, segmentation and classification methods in the state of the art (SVM: Support Vector Machines, P: Precision, R: Recall, OA: Overall accuracy). Colors indicate similar methods used by different authors.

Authors	Semantic analysis methods	Number of classes	Accuracy
Mallet et al. (2008)	Full–waveform analysis, Mathematical morphology, SVM	3 (buildings, ground, vegetation)	P=95.0%
Golovinskiy et al. (2009)	Elevation images, graphs, contextual analysis, hierarchical clustering, SVM	16 (cars, pole–like objects, trash cans, parking meters, among others)	P=58%, R=65%
Hernández and Marcotegui (2009c)	Elevation images, mathematical morphology, SVM, linear discriminant analysis	4 (cars, lampposts, pedestrians, other)	P=86.21%
Munoz et al. (2009)	Contextual analysis, clustering, high-order Markov models	5 (vegetation, wires, poles/trunks, load bearing, facades)	P=87.1%
Owechko et al. (2010)	3D strip by strip processing, decision trees	17 (Buildings, ground, cars, bollards, lampposts, trees, among others)	P=70.0%
Zhu et al. (2010)	Elevation images, graph–cuts, SVM, decision trees	7 (buildings, bushes, cars, trees, pedestrians, bicycles, other)	P=89.6%
Demantke et al. (2010)	3D adaptive neighborhood, principal component analysis, decision trees, dimensionality features	4 (lines, planes, volumes, noise)	P=69.3%
Douillard et al. (2011)	Voxelization, hierarchical clustering, decision trees, RANSAC, clustering	16 (ground and several urban objects)	P=89.0%
Rutzinger et al. (2011)	3D Hough transform, region growing, shape models, 3D alpha shapes	2 (trees, non-tree)	P=93%, R=86%
Pu et al. (2011)	Geometrical and topological analysis, decision trees	3 (poles, trees, other)	P=73.5%
Velizhev et al. (2012)	RANSAC, hierarchical clustering, spin images, implicit shape models	2 (cars, light poles)	P=69%, R=80%
Weinmann et al. (2013)	Elevation images, 3D adaptive neighborhood, SVM, k-nearest neighbors, Naive Bayesian	5 (wire, pole/trunk, facade, ground and vegetation)	OA=93.3%
Niemeyer et al. (2014)	3D point by point processing, random forests, Markov random fields	7 (grassland, road, ground roof, low vegetation, facade, flat roof and trees)	OA=83.4%

images. The 3D point cloud is projected from the sensor point of view. As a result, a "panoramic" range image is obtained and plane estimations are done for each pixel on the image. Then, a region growing approach is performed in order to segment pixels belonging to the same plane. In a similar way, Zhu et al. (2010) project Mobile Laser Scanning (MLS) data to a "panoramic" range image in which rows represent the acquisition time of each laser scan–line, columns represent the sequential order of measurement and pixel values code the distance from the sensor to the point. They propose a semantic analysis using graphs, SVM and decision trees. Hernández and Marcotegui (2009c) propose a method projecting MLS data to elevation images, *i.e.* a nadir view of the scene. Ground and objects are segmented using morphological transformations and objects are classified in four categories (cars, lampposts, pedestrians, and other) using SVM.

Since processing based on elevation images is both precise and fast, real-time applications such as automatic guided vehicles have been addressed. Kammel et al. (2008) and Ferguson et al. (2008) have developed autonomous vehicles, for the DARPA Challenge 2007, able to drive through urban scenarios. They use off-line processed aerial images and 2D maps in order to determine road structure. Then, on-line laser scans are projected to elevation images and static and mobile obstacles are detected. Munoz et al. (2009), extending the work by Anguelov et al. (2005), propose High Order Markov Random Fields for on–board contextual classification. In general, approaches for autonomous vehicles do not require high (centimeter) accuracy but high speed in order to detect and predict obstacles in real time. More accurate but slower methods process the 3D point cloud directly. These approaches are suitable for applications with high accuracy requirements but no strict time constraints. One of the major problems is the 3D neighborhood definition, which is not as trivial as it is in the 2D case using elevation images. Demantke et al. (2010) propose a method to adapt 3D neighborhood radius

based on local features. Radius selection is carried out optimizing local entropy. Then, dimensionality features are calculated on spherical neighborhoods in order to characterize lines (1D), planes (2D) and volumes (3D). Douillard et al. (2011) present a set of 3D segmentation methods based on voxelization and meshing. Their algorithms are evaluated on manually labeled datasets and the best performance is achieved using clustering approaches.

Several authors develop hybrid methods exploiting the complementarity between passive and active 3D acquisition methods: laser scanning provides the accurate 3D geometry while photogrammetry provides the realistic texture. Sevcik and Studnicka (2006) present a method based on laser scanning and photogrammetry for generating precise and detailed 3D city models. Beger et al. (2011) use both high resolution images and airborne LiDAR data to generate 3D orthophotos with depth information. Gerke and Xiao (2014) combine Aerial Laser Scanning (ALS) and images for automatic scene classification.

Several complete semantic analysis frameworks can be also found in the literature. Golovinskiy et al. (2009) develop a set of algorithms to detect, segment, characterize and classify urban objects. Their method is evaluated on ALS/TLS data from Ohio (USA). Their pipeline is as follows: i) ground segmentation using graph cuts; ii) object detection and segmentation using hierarchical clustering; iii) object characterization using geometrical and contextual descriptors; iv) object classification using SVM. Recently, Velizhev et al. (2012) have improved this work–flow including spin images and implicit shape models. The major problems of these approaches are noise, sparse sampling and proximity between objects. Moreover, some prior knowledge about the object scale is required to set up thresholds. Schnabel et al. (2008) present a semantic system for 3D shape detection. Their algorithm consists in two main steps: i) a topology graph is built with primitive shapes extracted from the data; ii) a search is carried out in order to detect characteristic subgraphs of semantic entities. The main drawback is the graph complexity when dealing with non-trivial objects. Weinmann et al. (2013) propose a methodology for feature relevance assessment on 3D urban data. They propose a metric based on seven different feature selection strategies. Their results reveal that the use of the five best-ranked features improves the classification accuracy and reduce processing time and memory consumption. Niemeyer et al. (2014) address the problem of contextual classification on ALS data. In the framework of that work, no segmentation is performed and each 3D point is classified to one of seven categories using random forests and conditional random fields. After a feature analysis, the authors concluded that geometrical features, in particular the relative height, and contextual features are the most discriminant. Pu et al. (2011) propose a framework to segment and classify urban objects from MLS data. That work starts with a rough classification into three large categories: ground, on-ground objects and off-ground objects. Then, based on geometrical attributes and topological relations, more detailed classes such as traffic signs, trees, building walls and barriers are recognized. Owechko et al. (2010) describe a similar pipeline: first, a spatial cueing is applied in order to identify potential objects; then, statistical classifiers based on decision trees are trained using geometrical and contextual features. Using such pipeline, there is barely any problem recognizing large flat features such as ground, barriers and walls. However, there are some problems classifying pole-like objects such as trees, bollards and lampposts. Additionally, occlusions and point density distribution are critical. Mallet et al. (2011) investigate the potential of full-waveform LiDAR data for urban areas classification. In that work, waveform features are used as input for an SVM classifier. Their results show that echo amplitude and radiometric features are suitable to classify buildings, ground and vegetation. Rutzinger et al. (2011) describe an automated work-flow to segment and to model trees from MLS data. First, the input point cloud is segmented into planar regions using the 3D Hough Transform and surface growing algorithms. Then, the remaining small segments are merged applying a connectivity analysis. Next, non-tree objects are removed from the analysis using statistical measures. Finally, trees are thinned using 3D alpha shapes (Edelsbrunner and Mücke, 1994) and realistic 3D models are generated. Zhou and Vosselman (2012) segment and model curbstones from ALS/MLS data. Their process is performed directly on the 3D point cloud, on a strip by strip basis, so intrinsic information between neighboring strips is missing. Recently, Serna and Marcotegui (2013b) solved this problem by processing all strips at the same time using elevation images.

6.4 Object detection

Our object detection method is based on mathematical morphology, inspired by Hernández and Marcotegui (2009a). They propose to detect urban objects using the top-hat by filling holes (THFH) followed by an area opening. In the first step, THFH is an effective and parameterless way to extract objects that appear as bumps on the elevation image. However, it fails extracting objects touching the image border because they are not considered as bumps. In the second step, an area opening $\gamma_{A_{min}}$ (Vincent, 1994) is performed in order to filter out small and noisy structures. Area opening is a morphological filter that removes objects with an area smaller

than a given threshold A_{min}. This procedure is effective to get rid of noisy and isolated regions. However, it also removes thin objects such as poles. In general, pole-like objects have a small area when they are seen from a nadir point of view, so they are suppressed by this filter. In this section, we propose an object detection framework that solves these two problems.

In order to solve the drawbacks of THFH step, a twofold strategy is proposed. A structure is considered to be object candidate if at least one of the two following conditions is fulfilled: i) it has not been reached by the quasi-flat zones algorithm (presented in Section 4.4), *i.e.* it does not belong to ground mask $\hat{f}_{gr}$; ii) it appears as a bump on interpolated elevation image $\hat{f}$ (presented in Section 3.6.3). Therefore, the first set of object candidates is the ground residue, which is computed by the arithmetic difference between the interpolated elevation image and the ground mask ($\hat{f} - \hat{f}_{gr}$). The second set of object candidates is extracted using transformation THFH($\hat{f}$), as originally proposed by Hernández and Marcotegui (2009a). Then, the union of these two sets constitutes the complete collection of object candidates.

In order to solve the $\gamma_{A_{min}}$ drawbacks, the normalized accumulation image $f_{\overline{acc}}$ is used (presented in Section 3.6.2). In general, vertical structures have high accumulation values. Thus, pole-like objects can be easily reinserted since their accumulation is higher than the accumulation for noisy objects.

Let us explain our detection method with an example. Figure 6.3 illustrates a typical acquisition profile. The urban profile contains the following urban objects enumerated from ① to ⑦: ① car, ② pedestrian, ③ noisy structure, ④ dog, ⑤ pedestrian, ⑥ house facade, and ⑦ chimney. Note that this is only an illustrative example in the 1D case. The process is performed on the entire 2.5D elevation image f.

The first step consists in interpolating occluded zones using a fill holes transformation, as explained in Section 3.6.3. Figure 6.3(a) presents interpolated profile $\hat{f}$. Using this transformation, each hole is filled with the minimal value surrounding the hole. For example, consider the hole in the left part, between objects ③ and ④. This hole is filled at the ground level because in 2.5D it is connected to ground pixels. Additionally, consider the holes in the left part, between objects ② and ③, and in the right part, between objects ⑤ and ⑥. These holes are also filled at the ground level even if the ground is not the minimal surrounding value in this 1D profile. We assume that these holes can be filled at that level because the ground is not occluded by pedestrians ② and ⑤ in the previous or in the following profiles.

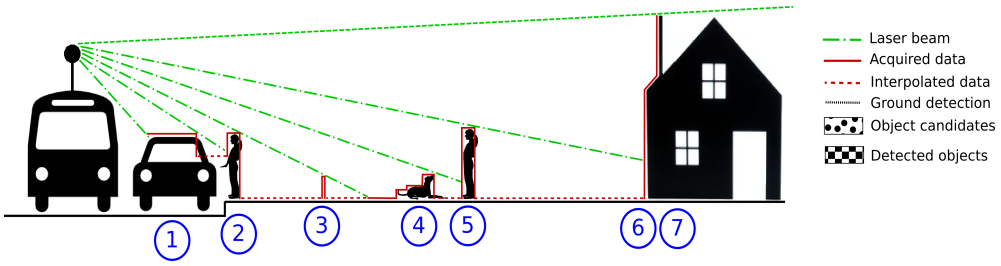
Figure 6.3(b) presents the first set of object candidates obtained as the ground residue ($\hat{f} - \hat{f}_{gr}$). Note that almost all objects are retrieved. However, the dog in the middle of the sidewalk (object ④) is not detected because it is too low. Thus, it has been reached by the quasi-flat zones propagation and it belongs to ground mask $\hat{f}_{gr}$.

In order to obtain the second set of object candidates, the profile is inverted and holes are filled using the morphological fill holes transformation, as shown in Figure 6.3(c). Then, transformation THFH($\hat{f}$)=Fill($\hat{f}'$)-$\hat{f}'$ consists in subtracting inverted image $\hat{f}'$ from inverted filled image Fill($\hat{f}'$), as shown in Figure 6.3(d). Note that this transformation detects correctly the dog in the middle of the sidewalk (object ④). However, the car in the left part (object ①) and the house in the right part (objects ⑥ and ⑦) are not retrieved because they are touching the border, then they do not become holes in the inverted profile. Figure 6.3(e) presents the complete set of object candidates, computed as the supremum between the two aforementioned sets of candidates: $(\hat{f} - \hat{f}_{gr}) \vee \text{THFH}(\hat{f})$.

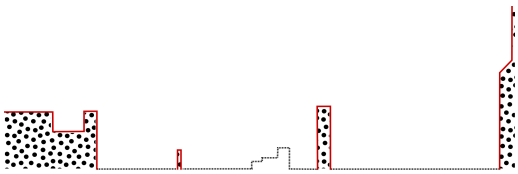
Figure 6.3(f) illustrates the effect of area opening $\gamma_{A_{min}}$ used to eliminate small and noisy structures. Note that the noisy structure in the middle of the sidewalk (object ③) has been correctly eliminated. However, the chimney (object ⑦) has also been suppressed. Finally, Figure 6.3(g) shows the result of the detection process, where the chimney has been reinserted because it has an important accumulation value.

Figure 6.4 illustrates the detection process on real data. Note that all objects are detected by our method. For a better understanding, facades are marked in a different color. In our experiments, facades are the highest vertical objects on the urban scene and they appear as elongated structures on interpolated maximal elevation image $\hat{f}$. Thus, they are segmented using morphological methods based on geometric and geodesic attributes. Facade segmentation methods have been introduced in Chapter 5.

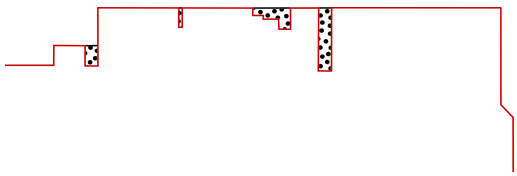
Figure 6.5(a) illustrates the pole-like object reinsertion. Note that several pole-like objects are removed by an area opening filter at A_{min}=0.1 m^2. In Figure 6.5(b), objects with an accumulation higher than 10 points are reinserted (in red). Note that a tilted bollard (black) is not recovered because it has not enough accumulation. A lower threshold can be used in order to retrieve this tilted bollard but at the risk of preserving other noisy structures.



(a) Acquisition scheme and interpolated profile $\hat{f}$.

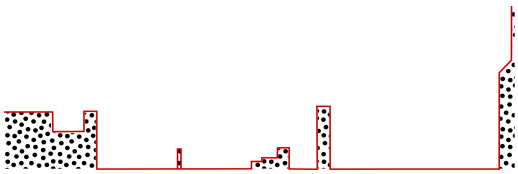


(b) In black: ground mask $\hat{f}_{gr}$. In red: first set of object candidates $\hat{f} - \hat{f}_{gr}$

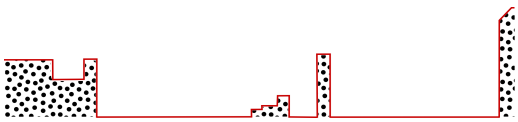


(c) Inverted profile $\hat{f}'$ and fill holes Fill($\hat{f}$')

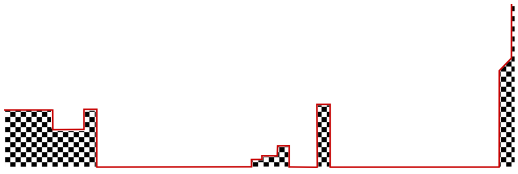
(d) Second set of object candidate: THFH(f)=Fill($\hat{f}'$) $- \hat{f}'$



(e) Complete set of object candidates: $(\hat{f} - \hat{f}_{gr}) \vee \text{THFH}(\hat{f})$

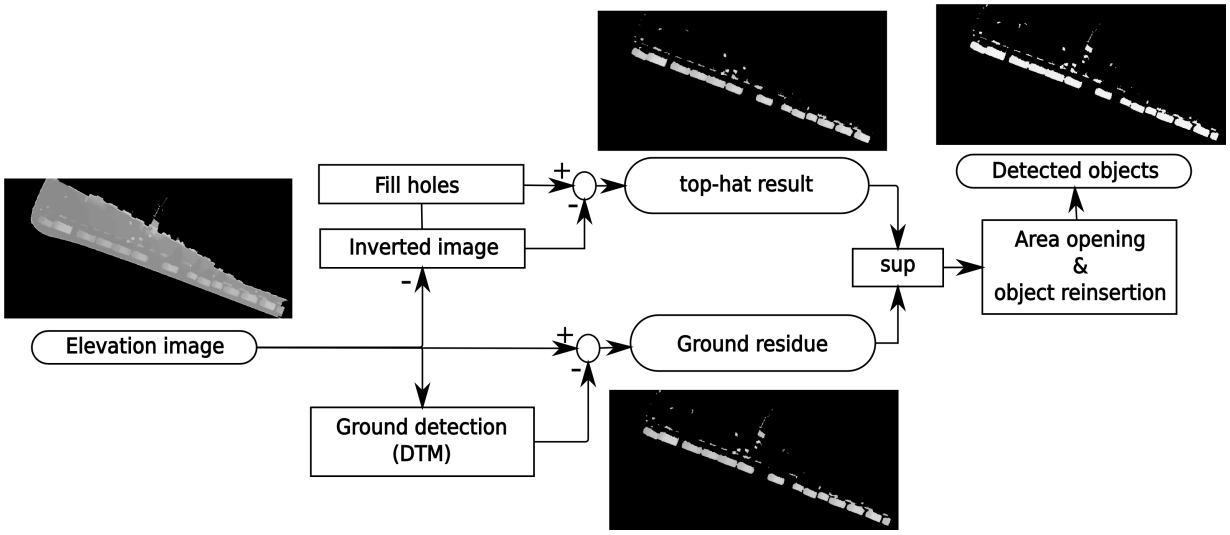


(f) Area opening: $\gamma_{A_{min}}((\hat{f} - \hat{f}_{gr}) \vee \text{THFH}(\hat{f}))$

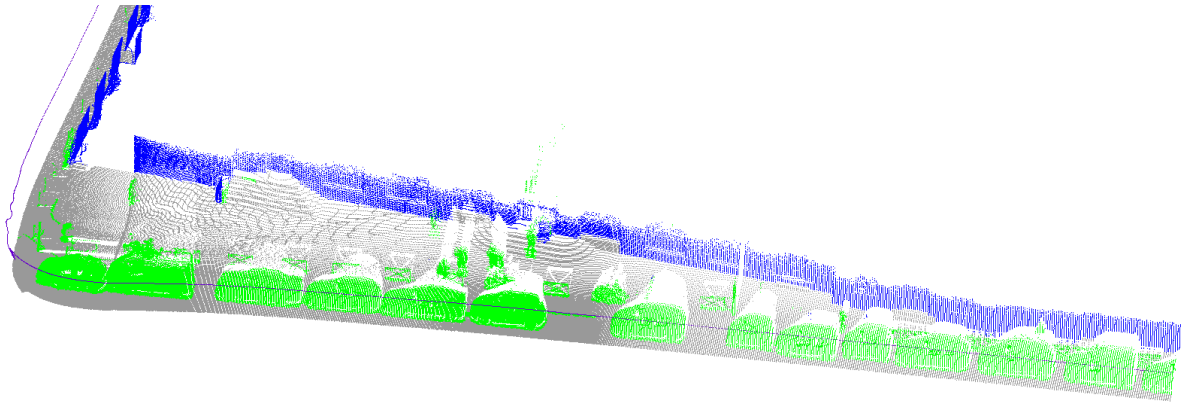


(g) Objects mask $\hat{f}_{obj}$ and ground mask $\hat{f}_{gr}$. Objects with significant accumulation $f_{\overline{acc}}$ are reinserted.

Figure 6.3: Detection method on a 1D profile. This urban scenario contains the following objects: ① car, ② pedestrian, ③ noisy structure, ④ dog, ⑤ pedestrian, ⑥ house facade, and ⑦ chimney.



(a) Detection scheme using elevation images



(b) Reprojection onto the 3D point cloud: ground (gray), objects (green), facade (blue) and acquisition trajectory (violet).

Figure 6.4: Object detection using the top-hat by filling holes and the ground residue. Note that all objects are detected by our method. For a better understanding, facades are marked in a different color. Facade segmentation methods have been introduced in Chapter 5.

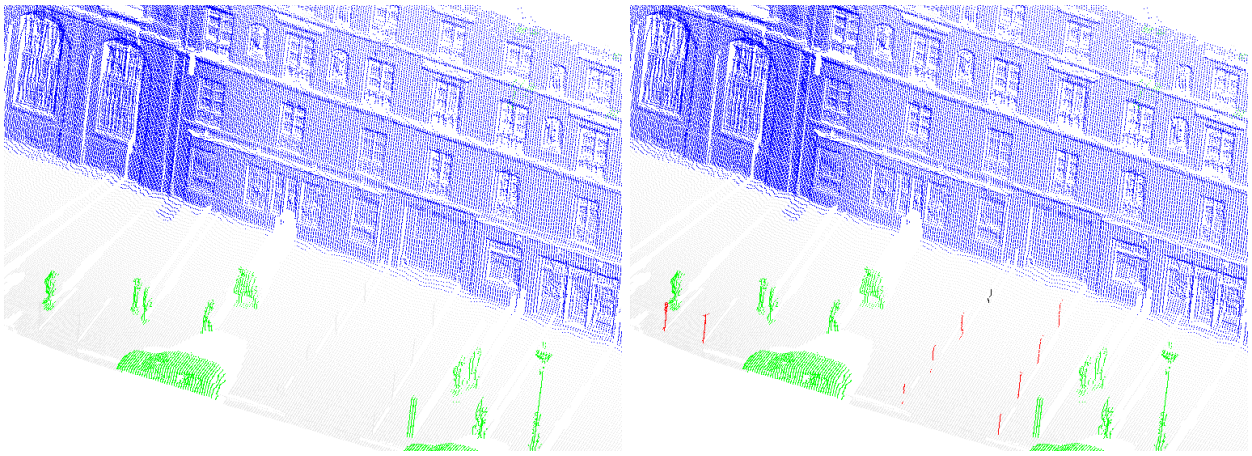
6.5 Object segmentation

One of the main drawbacks processing 3D urban data using elevation images is that high objects may occlude lower objects located below them. For example, in Figure 6.6, the pedestrian in the right part does not appear on the elevation image because it is below a tree. To solve this problem, we propose a segmentation strategy using two slices, as previously introduced in Section 3.5:

1. a lower slice, containing points between the ground level and a given height H_{slice} in the vertical axis. This slice is built to contain most urban objects.

2. an upper slice, containing points higher than H_{slice}. This slice contains the highest objects such as facades, treetops, lampposts and off-ground objects.

In our experiments, H_{slice} has been experimentally set to 3.5 m, which is usually high enough to include all obstacles for urban mobility. This slice separation is marked with a blue dotted line in Figure 6.6. This threshold can be modified in order to define obstacle maps at different heights according to different types of mobility: children, persons using a wheelchair, etc. Note that in the case of a non-horizontal or a non-flat surface, it is very important to segment the ground in order to adapt each slice to be parallel to the terrain. Methods used for ground segmentation have been previously discussed in Section 4.4.



(a) Detected objects after area opening.

(b) Reinserted poles (red) and a tilted bollard not retrieved (black).

Figure 6.5: Pole reinsertion using accumulation. One of 10 bollards has not been reinserted because it is tilted, thus it has no enough accumulation. (a) several pole-like objects are removed by an area opening filter at A_{min}=0.1 m^2. (b) objects with an accumulation higher than 10 points are reinserted (in red). Note that a tilted bollard (black) is not recovered because it has not enough accumulation. A lower threshold can be used in order to retrieve this tilted bollard but at the risk of preserving other noisy structures. Test site *rue Soufflot* in Paris. Acquired by IGN©France.

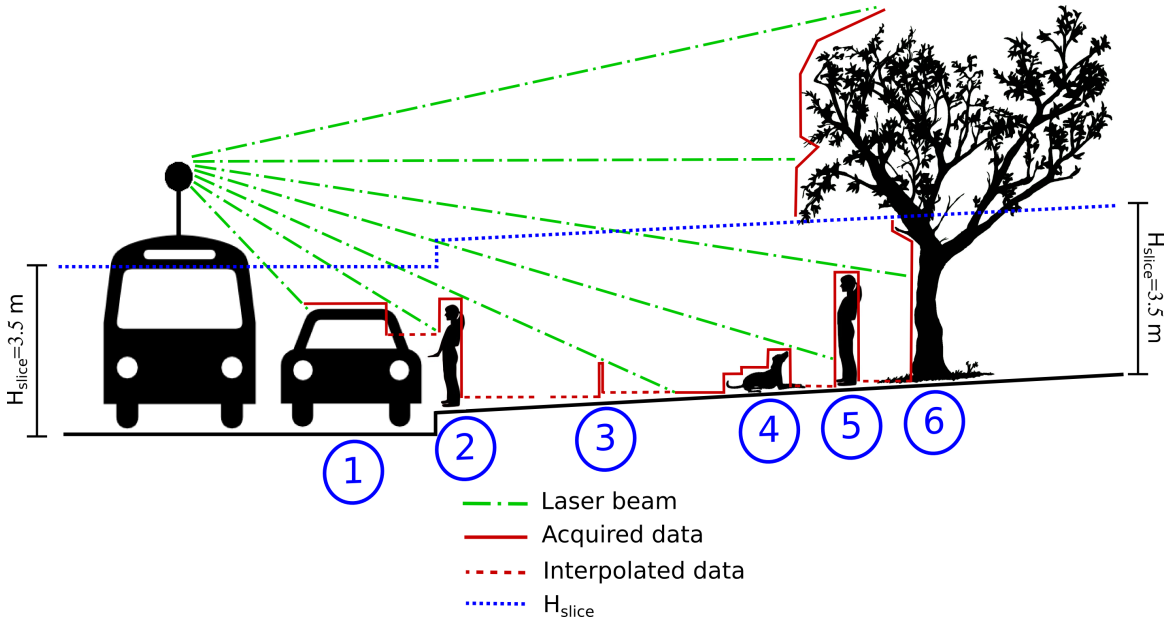
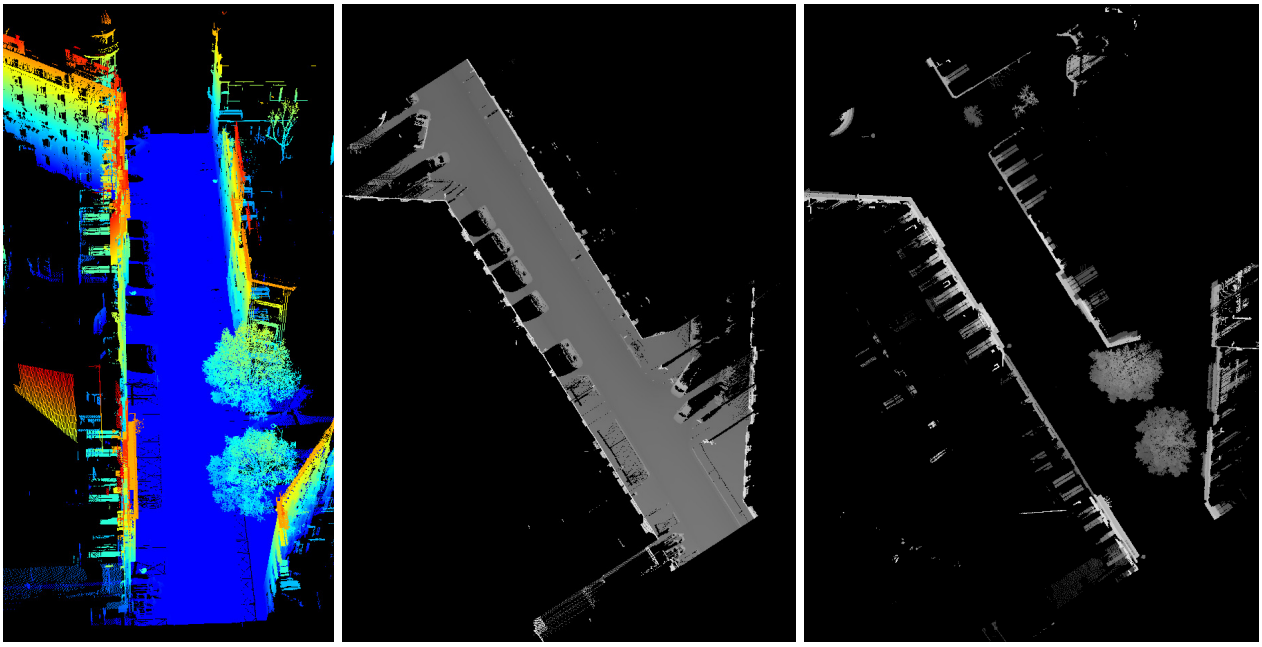


Figure 6.6: Slice definition in the 1D case. Note that processing by slices is useful to avoid that high objects such as trees (object ⑥) occlude lower objects below them such as pedestrians (object ⑤). In our experiments, H_{slice} has been experimentally set to 3.5 m, which is usually high enough to include all obstacles for urban mobility. This slice separation is marked with a blue dotted line. This threshold can be modified in order to define obstacle maps at different heights according to different types of mobility: children, persons using a wheelchair, etc. Note that in the case of a non-horizontal or a non-flat surface, it is very important to segment the ground in order to adapt each slice to be parallel to the terrain.

Figure 6.7(a) shows an experimental site in *rue d'Assas* in Paris and its corresponding lower (Figure 6.7(b)) and upper slices (Figure 6.7(c)). Note that trees and objects occluded below them can be processed separately on these two images. That is why this processing based on slices is particularly adapted to urban environments. After this slice definition, specific segmentation methods to analyze each slice have been developed, as explained later in Sections 6.5.1 and 6.5.2. Then, lower and upper results are integrated in order obtain coherent results, as presented in Section 6.5.3.



(a) 3D point cloud colored by the Z coordinate.

(b) Elevation image from the lower slice: points between the ground level and H_{slice}.

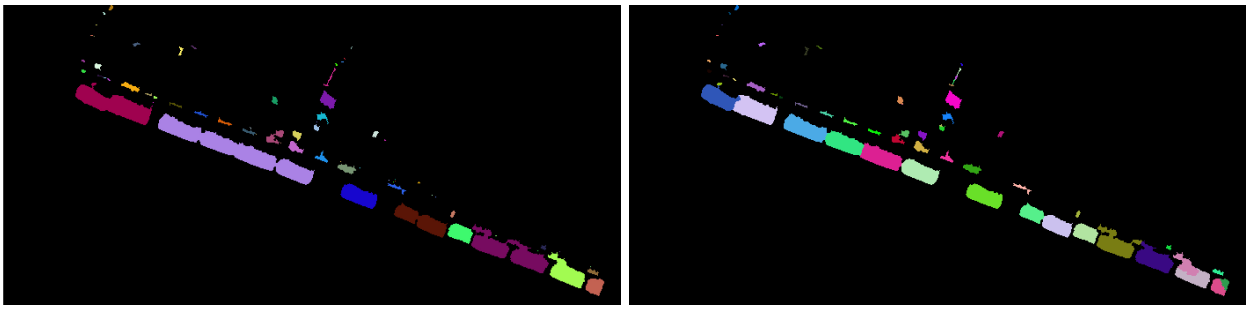
(c) Elevation image from the upper slice: points higher than H_{slice}.

Figure 6.7: Elevation images for lower and upper slices in a test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©. Note that trees and objects occluded below them can be processed separately on these two images. That is why this processing based on slices is particularly adapted to urban environments. After this slice definition, specific segmentation methods to analyze each slice have been developed, as explained later in Sections 6.5.1 and 6.5.2. Then, lower and upper results are integrated in order obtain coherent results, as presented in Section 6.5.3.

6.5.1 Object segmentation on the lower slice

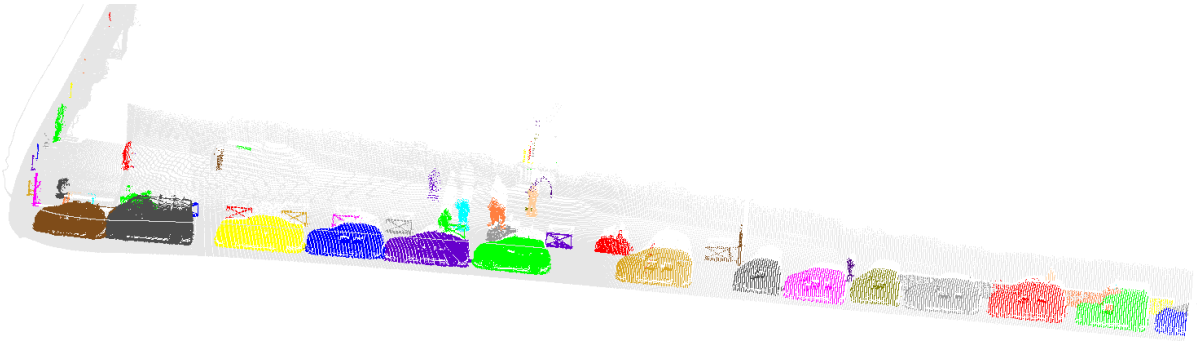
Using our detection approach (introduced in Section 6.4), it is possible to have several objects, close to each other, merged into a single CC. For example, in the left part of Figure 6.6, a car (object ①) and a pedestrian (object ②) are detected in the same CC. Another example is shown in Figure 6.8(a), where several cars are merged into a single CC. In order to solve this problem, we apply the solution proposed by Hernández and Marcotegui (2009c): "the number of connected objects in the same CC is equal to the number of significant maxima on it". With the aim of preserving only the most significant maxima, *i.e.* to get rid of maxima due to texture and noise on the upper part of the objects, a morphological h-Maxima filter is used (Schmitt and Preteux, 1986). The h-Maxima filter eliminates maxima whose relative height is less than or equal to a given threshold h, *i.e.* with a low local contrast. Using filtered maxima as markers, a constrained watershed on the elevation image is applied in order to segment connected objects. Figure 6.8 illustrates the performance of this segmentation.

The main drawback is when segmenting objects such as bikes, fences and lampposts with several arms. They may be over-segmented because they present more than one significant maximum on the elevation image. For overcoming this problem, shape and contextual information may help to decide whether an object should be re-segmented.



(a) Elevation image: connected objects

(b) Elevation image: segmented objects



(c) 3D point cloud: segmented objects

Figure 6.8: In order to segment connected objects we apply the solution proposed by Hernández and Marcotegui (2009c): "the number of connected objects in the same CC is equal to the number of significant maxima on it". With the aim of preserving only the most significant maxima, a morphological h-Maxima filter is used. The h-Maxima filter eliminates maxima whose relative height is less than or equal to a given threshold h, *i.e.* with a low local contrast. Using filtered maxima as markers, a constrained watershed on the elevation image is applied in order to segment connected objects. (c) illustrates the performance of this segmentation. Each color represents a different object. Test site in *rue Vaugirard* in Paris. IGN©France.

6.5.2 Object segmentation on the upper slice

Since the upper slice contains only the highest urban structures, we assume that only four kind of objects are found in this slice: facades, off-ground objects, trees and pole-like objects. Let us explain their segmentation process using the toy example of Figure 6.9. The 1D profile contains the following urban objects enumerated from ① to ⑤: ① facade, ② bird, ③ lamppost, ④ pedestrian, and ⑤ tree. Note that this is only an illustrative example in the 2D case, real process is performed on 2.5D elevation image f.

First, facades (object ①) are supposed to be previously segmented using one of the methods proposed in Chapter 5. Thus, they are extracted by simple comparison with the facade segmentation result.

Second, let us consider the case of objects which are not connected to the ground, as it is the case of the bird (object ②).

Definition 6.5.1 ***Off-ground object.*** *Let H_{slice} be the height at which a 3D point cloud is divided into two slices parallel to the ground, as proposed in Section 6.5. Let f^{up}_{min} and f^{up} be the minimal and maximal elevation images of the upper slice, respectively. By analogy, f^{low}_{min} and f^{low} stands for the minimal and maximal elevation images of the lower slice, respectively. An object X in the upper slice is an off-ground object if it is not connected to any object in the lower slice, which can be determined evaluating the elevation values at the slices boundary. Then, the set of off-ground objects $f_{off\text{-}gr}$ is defined as:*

$$f_{off\text{-}gr} = \{X \in f^{up} \mid min(f^{up}_{min}(X)) \neq H_{slice} \ \vee \ max(f^{low}(X)) \neq H_{slice}\} \tag{6.1}$$

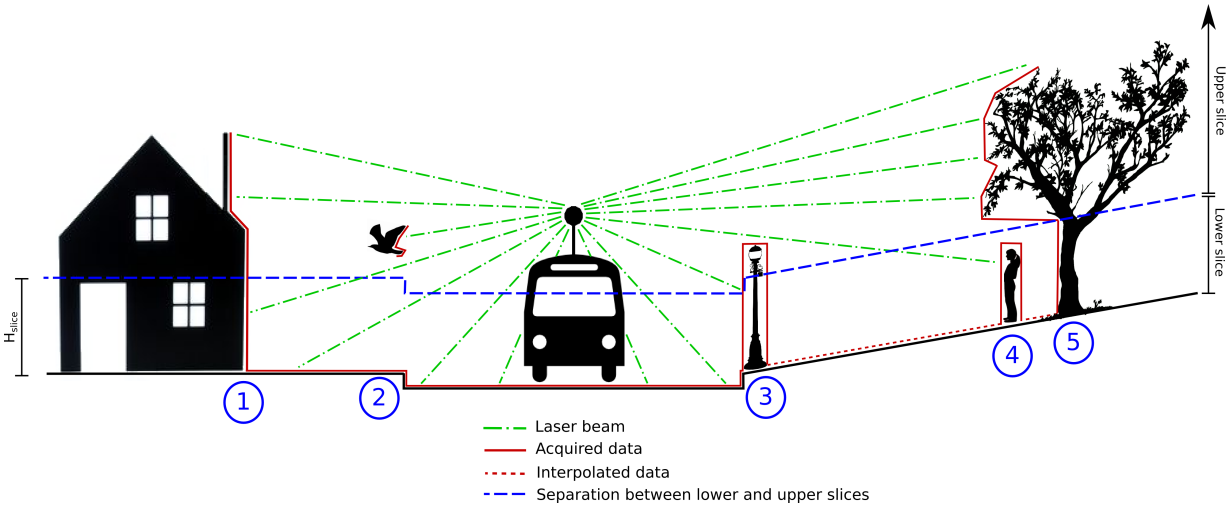


Figure 6.9: 1D example of object segmentation using two slices. This scenario contains the following objects: ① facade, ② bird, ③ lamppost, ④ pedestrian, and ⑤ tree. Note that this is only an illustrative example in the 2D case, real process is performed on 2.5D elevation image f.

Finally, trees and pole-like objects are the only remaining objects in the upper slice, as it is the case of objects ③ and ⑤. It is noteworthy that treetops are bigger than any pole-like object. Then, an area opening $\gamma_{A_{\text{tree}}}$ (Vincent, 1994) is used with the aim of segmenting trees and pole-like objects. Then, we define trees as objects bigger than a given threshold A_{tree}. Figure 6.10 shows an experimental zone with several trees in *St. Sulpice square* in Paris, France. Figure 6.10(a) presents the complete elevation image while Figure 6.10(b) presents its upper slice. Figures 6.10(c) to 6.10(f) present area thresholds (from 5 to 50 m^2) in order to segment trees. In our experiments, we have noted that A_{tree}=10 m^2 is enough to correctly segment trees while filtering out pole-like objects. However, this parameter can be intuitively tuned on any other database using some prior knowledge, *i.e.* type of pole-like objects, variety of trees, etc. In order to improve this segmentation process, features such as granulometry (Matheron, 1975), shape (Breen and Jones, 1996) or dimensionality attributes (Demantke et al., 2010) may be used.

6.5.3 Integrating lower and upper slices

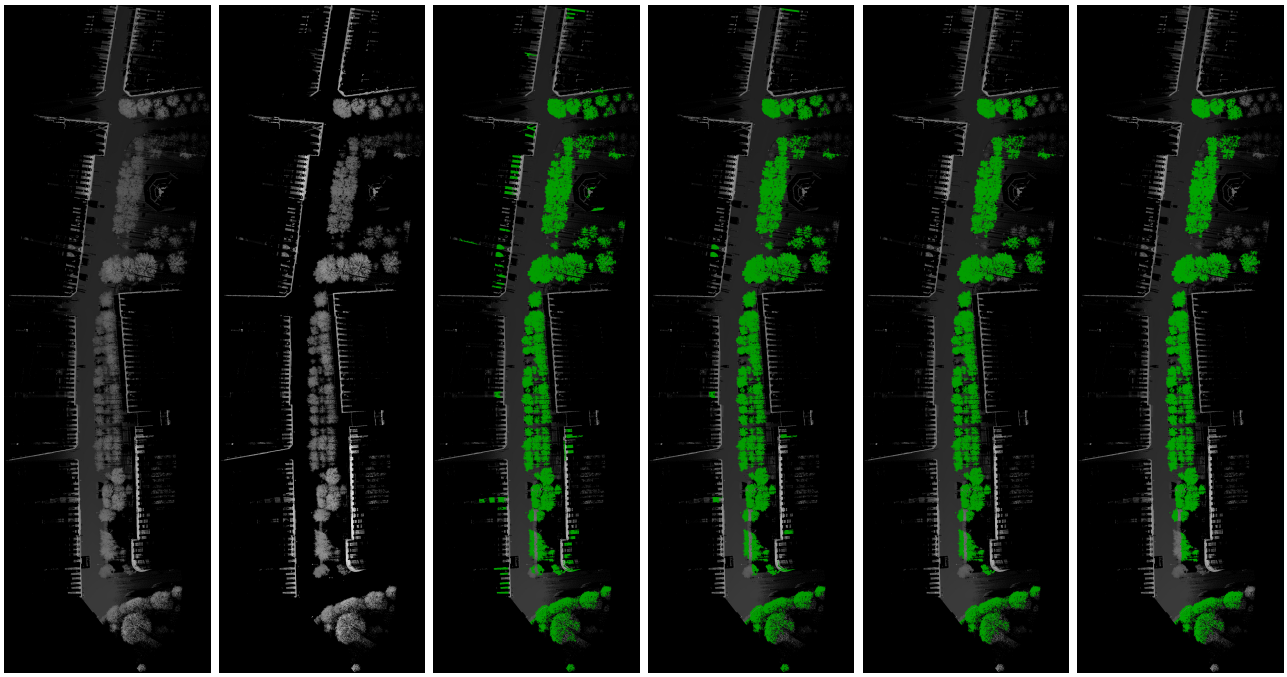
In order to obtain coherent results, lower and upper segmentation results should be integrated. As aforementioned, processing is independently carried out on each slice. Then, a connectivity should be defined in order to propagate results between slices.

Analyzing the elevation values at the slices boundary, as in Definition 6.5.1, it is possible to determine when an object in the upper slice is connected to another in the lower slice. Then, the propagation rules of Table 6.2 are applied to each CC.

Table 6.2: Propagation rules for results from lower and upper slices.

CC in the upper slice	CC in the lower slice	Procedure
Segmented as facade	indifferent	Upper and lower CC correspond to a facade and they should have the same *id.*
Segmented as off-ground	Not connected to the ground	Upper CC correspond to an off-ground object it should have a unique *id.*
Segmented as tree	Connected to the ground	Upper and lower CC correspond to a tree and they should have the same *id.*
Segmented as pole-like	Connected to the ground	Upper and lower CC correspond to a pole-like object and they should have the same *id.*

Another possible but slower solution could consider an adaptive voxelization, as that proposed in Section 5.5. Using such structure, 3D connectivity can be defined using 6- or 26-neighborhoods. Figure 6.11 illustrates an



(a) Complete elevation image. (b) Upper slice. (c) $A_{\text{tree}} > 5m^2$ (d) $A_{\text{tree}} > 10m^2$ (e) $A_{\text{tree}} > 20m^2$ (f) $A_{\text{tree}} > 50m^2$

Figure 6.10: Tree segmentation using different area thresholds. It is noteworthy that treetops are bigger than any pole-like object. Then, an area opening is used with the aim of segmenting trees and pole-like objects. Then, we define trees as objects bigger than a given threshold A_{tree}. (a) presents the complete elevation image while (b) presents its upper slice. (c, d, e and f) present area thresholds (from 5 to 50 m^2) in order to segment trees. In our experiments, we have noted that A_{tree}=10 m^2 is enough to correctly segment trees while filtering out pole-like objects. However, this parameter can be intuitively tuned on any other database using some prior knowledge, *i.e.* type of pole-like objects, variety of trees, etc. Test site in *St. Sulpice* square in Paris, France. Stereopolis II, IGN©. Second Experimental zone (Z2) in the TerraMobilita/iQmulus database.

adaptive voxelization using slices parallel to the ground.

6.6 Object classification

Several classification methods have already been applied to 3D data in urban areas. In general, supervised classifiers are preferred since they offer a higher performance. In addition to the feature vector, a set of labels associated to each training sample is required. This set is called the training dataset, which is used to estimate the parameters of the classifier. An important underlying assumption is that the whole dataset has similar feature distribution with respect to the training dataset. This means that test and training datasets must have similar features in order to achieve a good performance. To prevent over-fitting, several techniques such as bootstrapping or cross-validation can be used.

In our work, SVM is chosen because it has remarkable abilities to deal with both high-dimensional data and limited training sets, is easy to implement, uses a simple set of features as input, and produces accurate results in similar applications reported in the literature (Mallet et al., 2008; Hernández and Marcotegui, 2009c; Alexander et al., 2010; Mountrakis et al., 2011). Other methods such as random forests and high order Markov models could also be suitable and they are known for providing similar performance (Anguelov et al., 2005; Mallet et al., 2008; Munoz et al., 2009).

In order to build the feature vector, three set of features are used:

- **Geometrical features:** object area and perimeter; bounding box area; mean axes length; maximum,

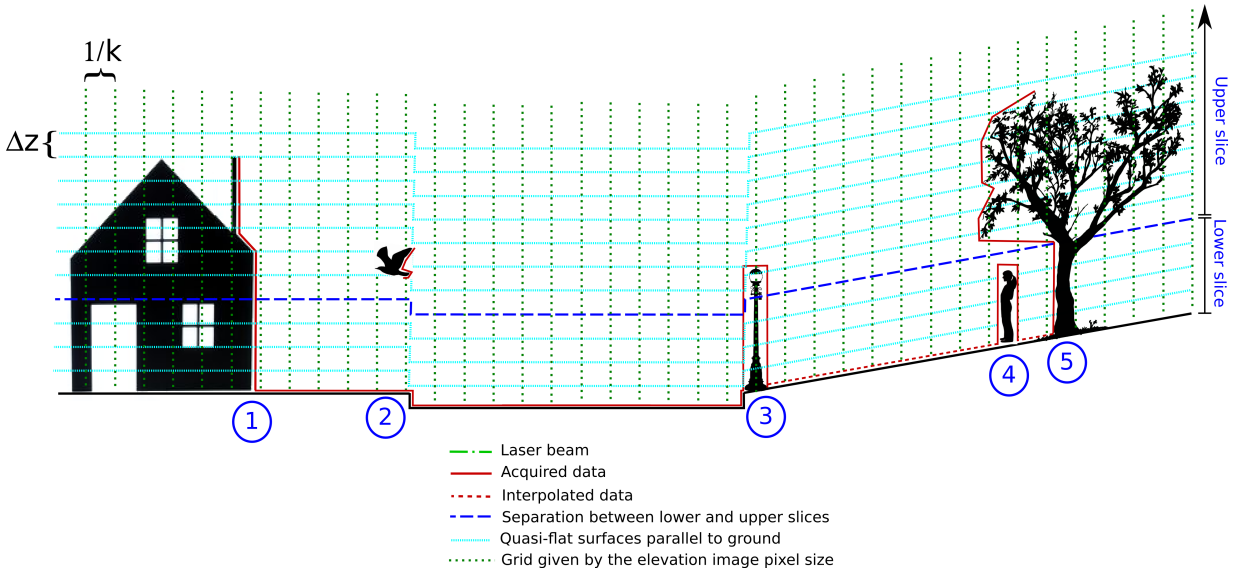


Figure 6.11: Adaptive voxelization using quasi-flat surfaces parallel to the ground. This example contains five objects: ① facade, ② bird, ③ lamppost, ④ pedestrian, and ⑤ tree. Using such structure, 3D connectivity can be defined using 6- or 26-neighborhoods.

mean, standard deviation and mode (the most frequent value) of the object height; object volume, computed as the integral of the height image over each object.

- **Contextual features:** Neighboring objects N_{neigh}, defined as the number of regions touching the object, using 8-connectivity on the elevation image. This feature is very discriminant in the case of group of trees and cars parked next to each other; confidence index $C_{ind} = n_{real}/(n_{real} + n_{interp})$, where n_{real} and n_{interp} are the number of non-empty object pixels before and after interpolation, respectively. In general, occluded and far objects have a low confidence index.

- **Intensity features:** Average laser intensity over the object. This feature is used if available.

- **Color features:** Average red, green and blue over the object. These features are used if available.

The reliability of these features depends on the acquisition system. Accurate and calibrated sensors contribute to compute accurate features and to get a good classification performance. Note that geometrical features can be adapted to any XYZ point cloud, taking into account the acquisition system resolution. In our experiments, geometrical features are computed in the international unit system (SI units).

6.6.1 Hierarchical classification

With the aim of reducing confusion between classes with similar features or with few examples in the database, we propose a hierarchical classification approach. Such idea comes directly from the study of biological perceptual systems (Hubel and Wiesel, 1962; Poggio and Shelton, 1999), and it has been also applied in the remote sensing community (Avci and Akyurek, 2004; Pu et al., 2011).

Our hierarchical classification begins using general classes, then it continues in a top-down approach until obtaining more detailed classes. First, data are separated in training and test subsets. The definition of the hierarchy of classifiers is entirely carried out on the training dataset. This approach can be implemented as follows:

1. An analysis is carried out on the training dataset applying a global classification taking all available classes into account;

2. Training errors are computed using a k-fold cross-validation approach. In k-fold cross-validation, we first divide the training set into k subsets of equal size. In our experiments, we have used k=10. Sequentially,

one subset is tested using the classifier trained on the remaining k-1 subsets. Thus, each instance of the whole training set is predicted once.

3. Classical Precision P(train), Recall R(train) and f_{mean}(train) = $(2 \times P(\text{train}) \times R(\text{train}))/(P(\text{train}) + R(\text{train}))$ statistics are computed in order to evaluate our training results. Classes with high confusion rates (f_{mean}(train) lower than 80%) are identified. In general, these classes correspond to heterogeneous objects with few examples. These classes are gathered in more general new classes.

4. Using the whole training dataset, two kind of classifiers are trained: the first one is a classifier trained with the well-distinguished original classes and the new general ones; the second one is a more specific classifier used for each new general class aiming at obtaining more detailed classes.

5. The process can be iterated. In our experiments, only two levels of hierarchy have been used.

After training, the resulting classifier is used to predict the test dataset. Precision $P(test)$, Recall $R(test)$ and $f_{mean}(test)$ results reported in Section 6.8 have been computed on the test dataset and reflect the performances of our system on real operation conditions.

6.7 TerraMobilita/iQmulus evaluation protocol

In order to benchmark our methods, we have cooperated with the National French Mapping Agency (IGN) in the definition of an evaluation protocol in the framework of TerraMobilita/iQmulus benchmark (Brédif et al., 2014). We propose a very detailed semantic tree containing 101 classes, shown in Figure 6.12. Probably no existing method in the state of the art treats the whole problem. This is why the participants to the benchmark can choose to analyze the scene using any subtree of the tree. In this case, they simply apply the *other* class to the nodes that they do not wish to detail. The evaluation is performed accordingly and only the relevant metrics are given.

The benchmark does not aim at ranking the participants but at providing insights on the strengths and weaknesses of each method. We consider that the quality of a method is subjective and application dependent, and the results of this benchmark should only help a user choosing one approach depending on its own specific requirements. Quality of the results is evaluated at three levels: classification, detection and segmentation. Details are given below.

6.7.1 Classification quality

The classification quality is evaluated point-wise . The result of the evaluation is a confusion matrix for each node of the tree. Rows and lines are the classes from the ground truth (GT) and the evaluated method, respectively. Matrix values are the percentage of points for each corresponding class. All nodes from the semantic tree have an *other* class, so participants can classify into less classes than those given in the tree. For non root nodes, an additional category *not in class* is given for each point that were not correctly classified at a lower level.

6.7.2 Segmentation quality

The segmentation quality measures the capacity of the method to retrieve the objects present in the scene. Thus, it requires to choose a criterion to determine if an object from the GT has been correctly segmented or not. This biases the evaluation as this choice will impact the result. The proposed solution is to give the evaluation result for a varying threshold m on the minimum object overlap. In the benchmark, an object is defined by the subset of points with the same object identifier. For a such subset S^{GT} of the ground truth and S^{AR} of the evaluated algorithm result, we validate S^{AR} to be a correct segmentation of S^{GT} (a match) iff:

$$\frac{|S^{GT}|}{|S^{GT} \cup S^{AR}|} > m \quad \text{and} \quad \frac{|S^{AR}|}{|S^{GT} \cup S^{AR}|} > m \tag{6.2}$$

where $|\cdot|$ denotes the cardinal (number of points) of a set. The standard Precision (P), Recall (R) and f_{mean} are then functions of m, as shown in Equations (6.3) to (6.5):

$$P(m) = \frac{\text{number of segmented objects matched}}{\text{number of segmented objects}} \tag{6.3}$$

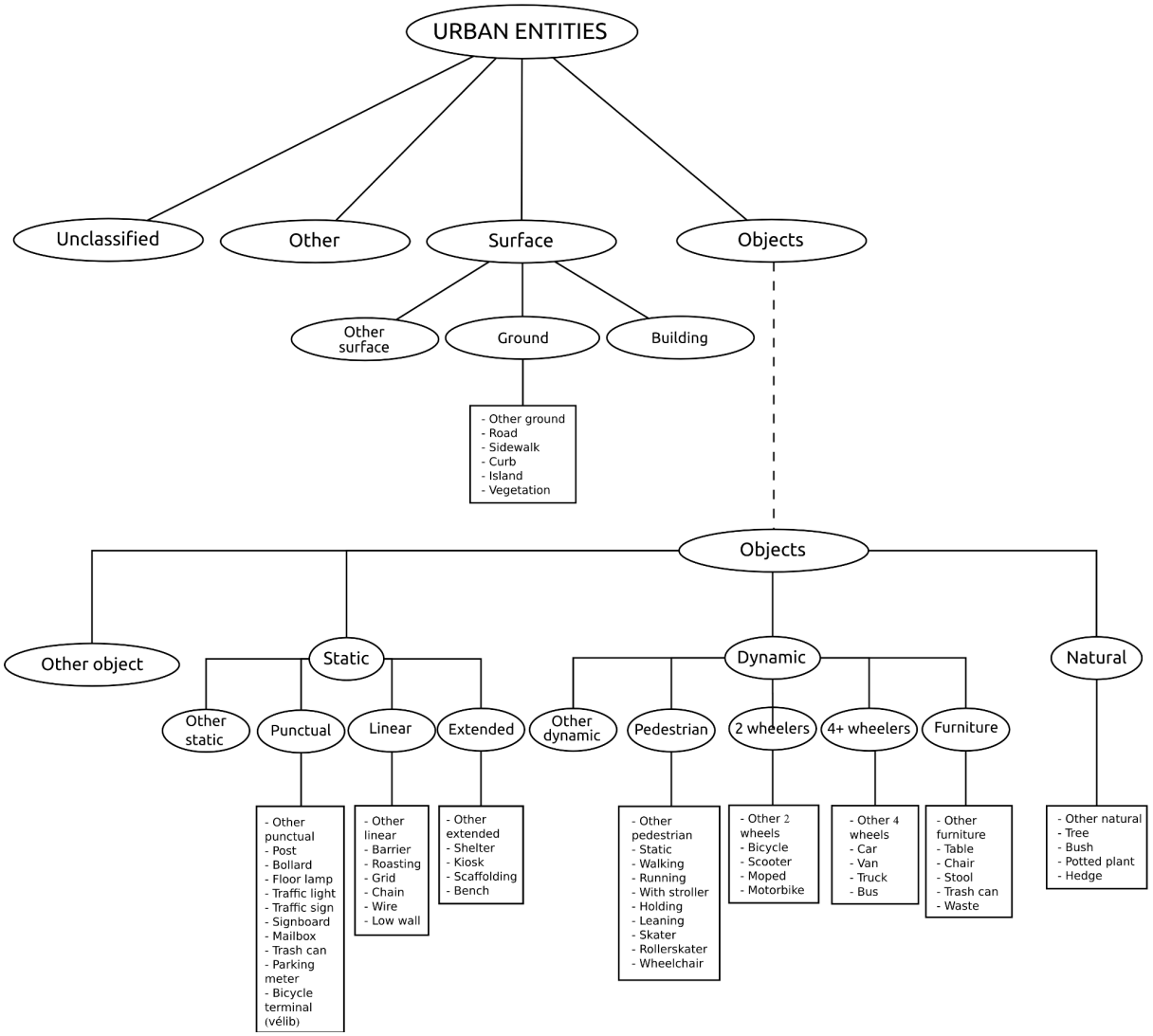


Figure 6.12: Hierarchy of semantic classes defined in the TerraMobilita/iQmulus benchmark. This class tree can be downloaded as an xml file from: http://data.ign.fr/benchmarks/UrbanAnalysis/download/classes.xml. We propose a very detailed semantic tree containing 101 classes. Probably no existing method in the state of the art treats the whole problem. This is why the participants to the benchmark can choose to analyze the scene using any subtree of the tree. In this case, they simply apply the *other* class to the nodes that they do not wish to detail. The evaluation is performed accordingly and only the relevant metrics are given.

$$R(m) = \frac{\text{number of segmented objects matched}}{\text{number of GT objects}} \tag{6.4}$$

$$f_{\text{mean}}(m) = \frac{2 \times P(m) \times R(m)}{P(m) + R(m)} \tag{6.5}$$

$P(m)$, $R(m)$ and $f_{\text{mean}}(m)$ are evaluated for each object type at each level of the semantic tree and results are presented as two curves. These statistics are decreasing in m and this decay indicates the geometric quality of the segmentation: the slower the decay, the better the segmented quality.

When the threshold m is below 0.5, criterion (6.2) does not guarantee that objects are uniquely matched. When $m < 1/n$, n objects from the GT can be matched to a single object of the algorithm result (AR), or

the opposite. Thus, for $m < 0.5$ we also give the curves of over-segmentation (1-to-n) and under-segmentation (n-to-1) by averaging n over the matches defined in Equation (6.2). These curves indicate the topological quality of the segmentation.

6.8 Results

Our methodology is evaluated on four databases: TerraMobilita/iQmulus (Section 6.8.1), Paris-rue-Soufflot (Section 6.8.2), Ohio (Section 6.8.3) and Paris-rues-Vaugirard-Madame (Section 6.8.4) databases. As a general remark, our experiments demonstrate that almost all objects are retrieved by our detection approach. Then, segmentation is useful to separate connected objects such as pedestrians and cars. However, bikes and bushes may be over-segmented. Finally, classification is carried out in an straightforward but effective way using an SVM approach with geometrical and contextual features. In our experiments, spatial pixel size ranges from 0.01 m^2 (pw=10 cm width) to 0.04 m^2 (pw=20 cm width) according to acquisition conditions.

It is noteworthy that our algorithms were initially developed to process 3D databases from Paris (France) in the framework of TerraMobilita project. One of the main advantages of our method is that it can be easily generalized to other datasets without any major modification. This is underlined by the good results obtained on Ohio database (Section 6.8.3). Detailed results are presented below.

6.8.1 Results: TerraMobilita/iQmulus database

TerraMobilita/iQmulus database (Brédif et al., 2014) has been developed aiming at benchmarking semantic analysis methods working on 3D dense urban data. This database has been created in the framework of TerraMobilita project. It consists in 11 annotated 3D point clouds acquired by Stereopolis II system in the 6^{th} Parisian district in January 2013. Annotation has been carried out in a manually assisted way by MATIS laboratory at IGN. Further details on this database can be found in Section 2.6.2.

For this experiment, the file "Cassette_idclass.ply" has been used[1]. It contains 12 million points from a street section approximately 200 m long in *rue Cassette* in Paris, France. Manual annotations and point-wise evaluations have been independently carried out by IGN, using the evaluation protocol presented in Section 6.7.

Figure 6.13 illustrates results for each step of our processing on the lower slice. Figure 6.13(a) presents the interpolated elevation image. Figure 6.13(b) shows the object detection result making the separation between ground, facades and objects. Note that objects are not individualized yet. Figures 6.13(c) and 6.13(d) present our segmentation and classification results, respectively. Objects with the same label must have the same class. Note that the main drawback is due to facades lower than H_{slice} since most of them are wrongly segmented as objects and classified as cars (zone D in Figure 6.13). Other problems are due to wrongly interpolated regions behind facades, which are segmented as objects (zones A, B and C in Figure 6.13). However, this is not so critical since they can be easily eliminated using the confidence index proposed in Section 6.6.

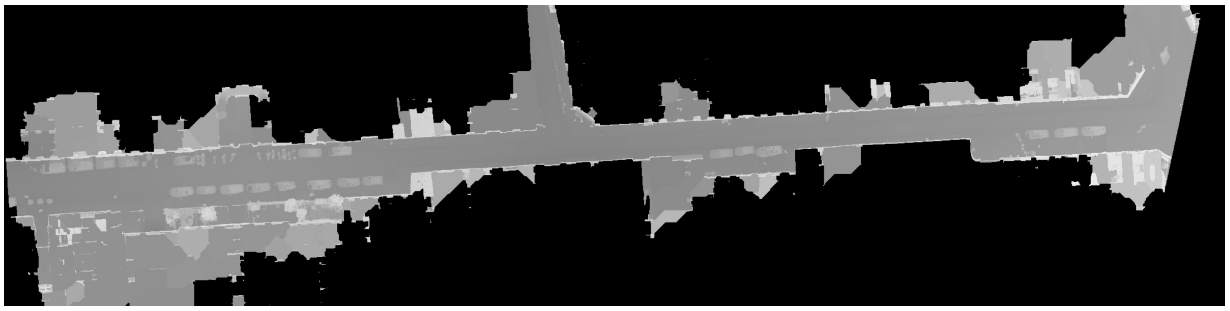
Figure 6.14 illustrates our segmentation and classification results on the upper slice. Figure 6.14(a) shows the elevation image. It is noteworthy that the upper slice only contains the highest urban structures such as facades, trees, poles and off-ground objects. Figures 6.14(b) and 6.14(c) show segmentation and classification results, respectively.

As aforementioned, TerraMobilita/iQmulus evaluation protocol has been used to evaluate our results (Brédif et al., 2014). First, the 3D point cloud is classified in 3 main categories: *surface* (containing facades and ground), *object* and *other*. Moreover, *unclassified* category is defined for non-annotated points in the GT. They are ambiguous points difficult to annotate, *e.g.* points behind facades, which correspond to 18.31 % of total number of points in the dataset. For example, consider the tree alignment in zone E in Figure 6.14. The tree in the left part touch a low facade below it. In the GT, several points of this tree have been manually marked as *unclassified*. These points have not been taken into account in the evaluation.

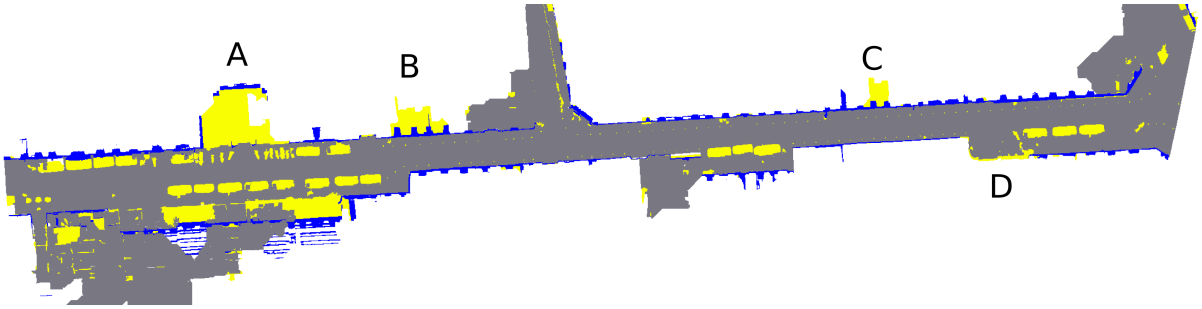
Table 6.3 presents the confusion matrix and our classification results for these 3 categories.

Using our method, *objects* are correctly detected with f_{mean} equal to 84.59 %. In this experiment, we are mainly interested in separating objects from other structures such as ground and facades. Note that *surface* class includes facades and ground, which represents the largest category in the scene with 75.82 % of total number of 3D points, while *object* class represents 5.7 % of total number of 3D points. The f_{mean} for *surface*

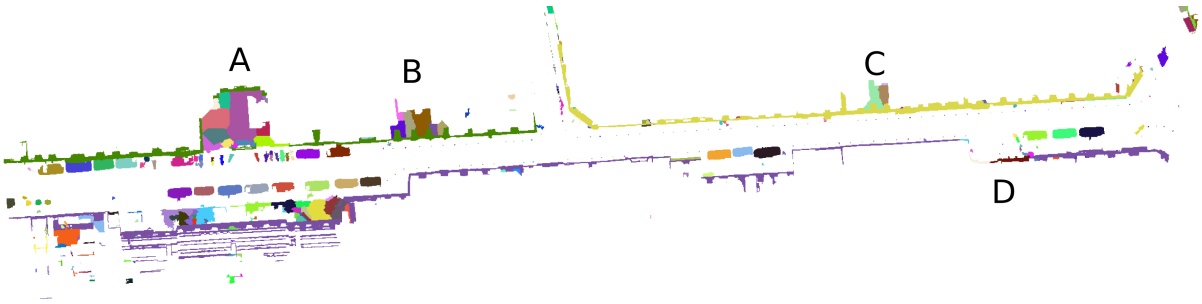
[1] The manual annotated 3D point cloud is available at:
http://data.ign.fr/benchmarks/UrbanAnalysis/download/Cassette_idclass.zip
The 3D point cloud processed by our method is available at:
https://partage.mines-telecom.fr/public.php?service=files&t=294aed38d48c8ddd03a528069f1b2e51



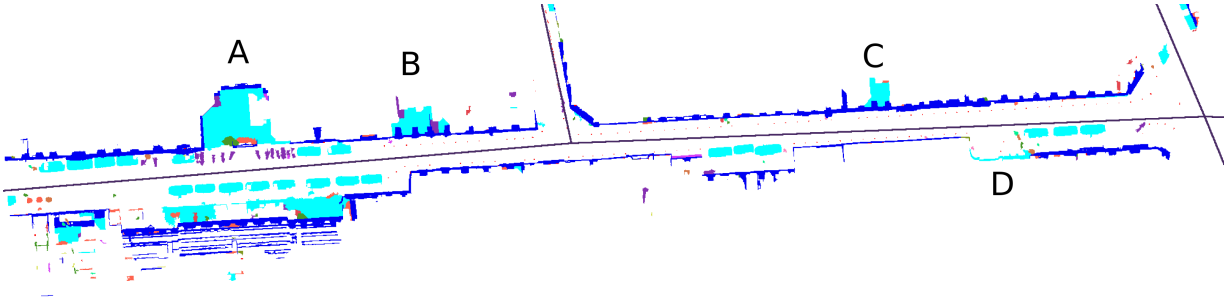
(a) Interpolated elevation image.



(b) Object detection result: ground (gray), facades (blue) and objects (yellow).

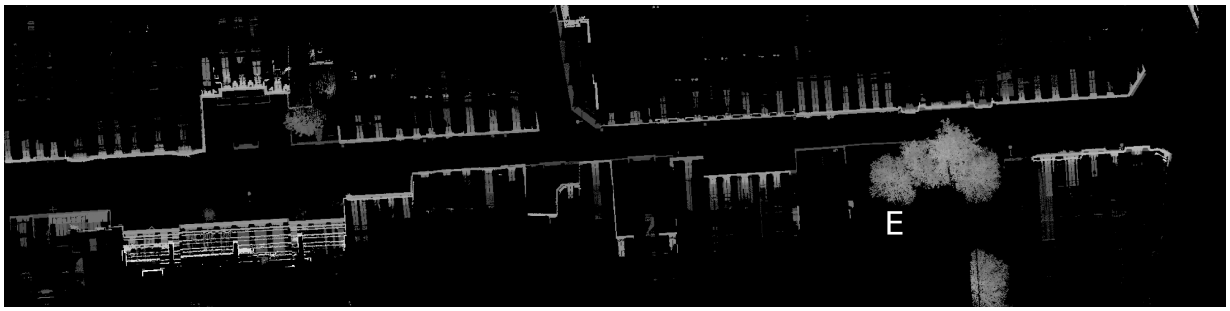


(c) Object segmentation: each color represents a different object.

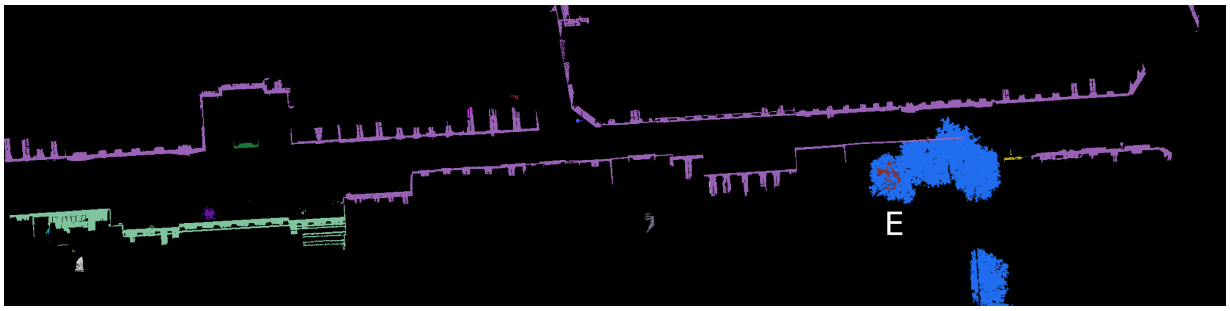


(d) Object classification: facades (blue), cars (cyan), bollards (red), motorcycles (indigo), pedestrians (orange), road medial axes (magenta).

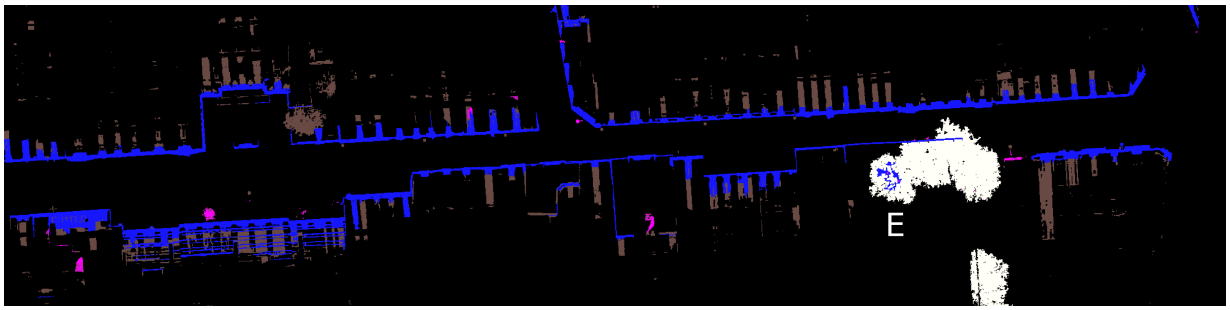
Figure 6.13: Semantic analysis on the lower slice on the "Cassette_idclass.ply" file. (a) interpolated elevation image. (b) object detection. Note that objects are not individualized yet. (c and d) present our segmentation and classification results, respectively. Objects with the same label must have the same class. Note that the main drawback is due to facades lower than H_{slice} wrongly segmented as objects and classified as cars (zone D). Other problems are due to wrongly interpolated regions behind facades, which are segmented as objects (zones A, B and C). Input file taken from TerraMobilita/iQmulus database. Acquired by IGN©France.



(a) Elevation image.



(b) Object segmentation: each color represents a different object.



(c) Object classification: facades (blue), trees (white), off-ground objects (brown), other (magenta).

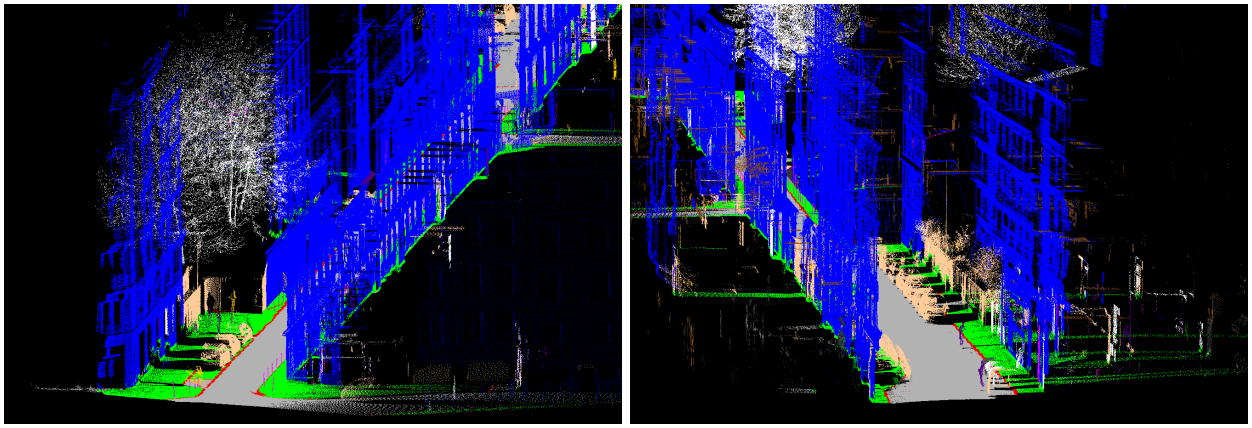
Figure 6.14: Semantic analysis on the upper slice on the "Cassette_idclass.ply" file. It is noteworthy that this slice only contains the highest urban structures such as facades, trees, poles and off-ground objects. Input file taken from TerraMobilita/iQmulus database. Acquired by IGN©France.

Table 6.3: Classification results for 3 general categories on TerraMobilita/iQmulus database. GT: ground truth, AR: automatic result. In the confusion matrix, results are presented as percentages with respect to the total number of points in the 3D point cloud (12 million points).

GT/AR	**unclassified**	**other**	**surface**	**object**	**Sum**	Recall	Precision	f_{mean}
unclassified	-	-	-	-	18.31 %	-	-	-
other	0.00 %	0.00 %	0.13 %	0.04 %	0.17 %	0.59 %	0.05 %	0.08 %
surface	1.90 %	2.19 %	70.81 %	0.91 %	75.82 %	93.40 %	98.82 %	96.03 %
object	0.09 %	0.02 %	0.72 %	4.88 %	5.70 %	85.49 %	83.72 %	84.59 %
Sum	1.99 %	2.21 %	71.66 %	5.82 %	81.69 %	**Overall accuracy:**		**92.65 %**

class is equal to 96.03 %. Figures 6.15(a) and 6.15(b) show two typical classification errors due to low facades wrongly detected as objects, usually cars. *Other* class is not correctly classified by our method, however it is not critical since it only represents 0.17 % of total number of 3D points in the scene. As aforementioned, *unclassified* class is not taken into account in the evaluation. Nevertheless, it is not critical in the practical case since it mainly contains 3D points behind facades, therefore they do not belong to the public space. The overall

accuracy of our method classifying these 4 main categories is 92.65 %.



(a) The low facade in the left part below the tree has been wrongly classified as object.

(b) The low facade in the right part behind cars has been wrongly classified as object.

Figure 6.15: Classification errors on TerraMobilita/iQmulus database. Facades (blue), sidewalk (green), road (gray), cars (pink), bollards (magenta), trees (white), pedestrians (indigo). Test zone in *rue Cassette* in Paris, France. Stereopolis II, IGN©.

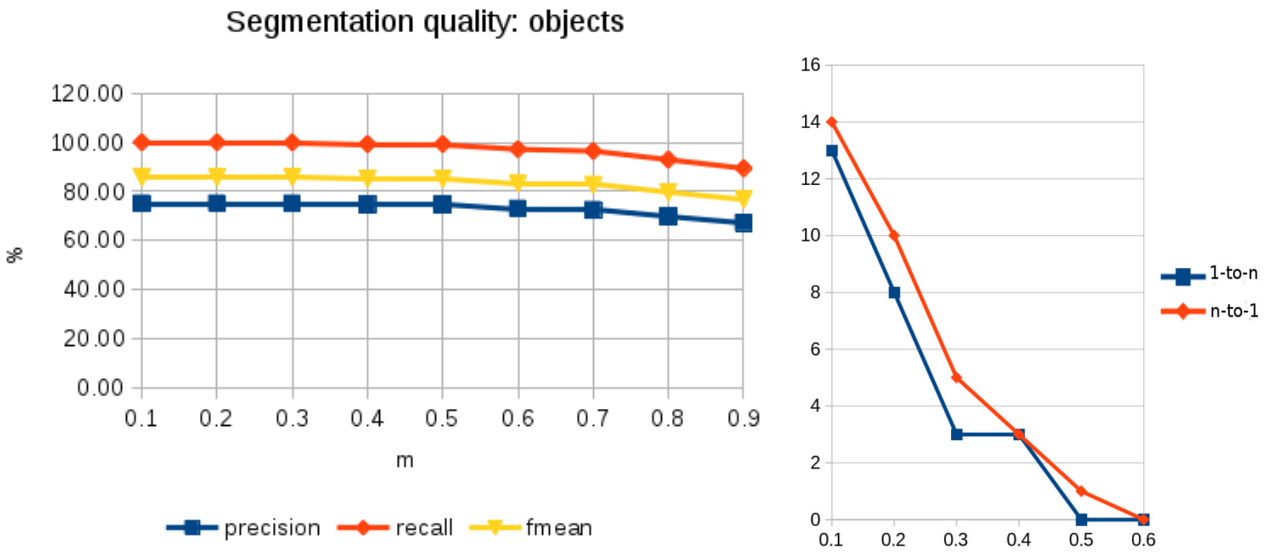
In order to evaluate the segmentation quality for *object* class, we use a varying threshold m defining the minimum object overlap to validate a segmentation (as explained in Section 6.7.2). P(m), R(m) and $f_{\text{mean}}(m)$ are evaluated for each m value and results are presented in Figure 6.16(a). These functions are decreasing as m and their decay indicate the geometric quality of the segmentation. The total number of objects annotated in the GT is 189. According to our results, our segmentation method retrieves 142 objects (P(0.1)=75.13 %) for m=0.1, while 127 object are retrieved (P(0.9)=67.20 %) for m=0.9. The geometric quality of our segmentation is good since the performance decays slowly. For example, $f_{\text{mean}}(m)$ decays from 85.80 % to 76.74 % for m varying from 0.1 to 0.9. Note that in the range m=[0.1, 0.5] our performances are constant while in the range m=[0.5, 0.9] the $f_{\text{mean}}(m)$ decays less than 10%, proving the robustness of our segmentation.

The topological errors of the segmentation for *object* class is given in Figure 6.16(b). The (1-to-n) and (n-to-1) curves indicate the over-segmentation and under-segmentation errors, respectively. They depend on threshold m used for matching, as explained in Section 6.7.1. Low thresholds induce high topological errors (both under- and over-segmentation). A threshold $m = 0.5$ is a good compromise for this method since precision/recall stay high (Figure 6.16(a)) while topological errors are not allowed (Equation (6.2)).

Table 6.4 shows classification results for *objects* subtree considering 3 categories: *static*, *dynamic* and *natural*. Using our method, *dynamic* and *natural* objects are correctly classified with f_{mean} equal to 92.56 % and 95.49 %, respectively. Note that *dynamic* and *natural* classes represent the largest structures in the scene with 92.90 % of all 3D points in the subtree. The main drawback is due to *static* objects wrongly classified as *dynamic*, as it is the case of fences, barriers and low walls classified as cars. Other small errors are due to parking meters wrongly classified as pedestrians and bushes wrongly classified as cars. However, these errors are not critical since they represent only 7.10 % of total number of points in the subtree. The overall accuracy of our method classifying *object* subtree is 91.84 %.

Table 6.4: Classification results for *object* subtree on TerraMobilita/iQmulus database. GT: ground truth, AR: automatic result. In the confusion matrix, results are presented as percentages with respect to the total number of points in the 3D point cloud (12 million points).

GT/AR	**unclassified**	**static**	**dynamic**	**natural**	**Sum**	Recall	Precision	f_{mean}
unclassified	-	-	-	-	0.0 %	-	-	-
static	0.0 %	0.75 %	6.09 %	0.26 %	7.10 %	10.58 %	50.85 %	17.51 %
dynamic	0.15 %	0.72 %	83.63 %	0.85 %	85.36 %	97.98 %	93.13 %	95.49 %
natural	0.0 %	0.0 %	0.08 %	7.45 %	7.54 %	98.86 %	87.01 %	92.56 %
Sum	0.15 %	1.48 %	89.80 %	8.57 %	100.0 %	**Overall accuracy:**		**91.84 %**



(a) Segmentation quality for *object* class.

(b) Topological errors for *object* class.

Figure 6.16: Segmentation quality and topological errors for *object* class on the TerraMobilita/iQmulus database. The (1-to-n) and (n-to-1) curves indicate the over-segmentation and under-segmentation errors, respectively. They depend on threshold m used for matching. Low thresholds induce high topological errors (both under- and over-segmentation). A threshold $m = 0.5$ is a good compromise for this method since precision/recall stay high (a) while topological errors are not allowed (b).

Figure 6.17 presents the segmentation quality for *dynamic object* class. The total number of *dynamic objects* annotated in the GT is 113. According to the results, our segmentation method correctly retrieves 97 dynamic objects (P(0.1)=85.84 %) for m=0.1, while 88 dynamic objects (P(0.9)=77.88 %) are retrieved for m=0.9. The geometric quality of our segmentation is good since the performance decays slowly. For example, $f_{\text{mean}}(m)$ decays from 92.38 % to 83.81 % for m varying from 0.1 to 0.9. Note that in the range m=[0.1, 0.5] our performances are constant, while in the range m=[0.5, 0.9] the $f_{\text{mean}}(m)$ decays less than 9%, which proves the robustness of our segmentation process.

For *static objects* subtree, 99.26 % of them correspond to punctual objects such as bollards, posts and traffic lights. The poles reinsertion method proposed in Section 6.4 is particularly effective to retrieve this kind of objects. In our experiments, a f_{mean} equal to 99.63 % is reported, proving the performance of our approach. Figure 6.18 presents the segmentation quality for *static objects* node. The geometric quality of our segmentation is good since the performance decays slowly. For example, $f_{\text{mean}}(m)$ decays from 87.80 % to 78.05 % for m varying from 0.1 to 0.9. Note that in the range m=[0.1, 0.5] our performances are constant, while in the range m=[0.5, 0.9] the $f_{\text{mean}}(m)$ decays less than 10%, proving the robustness of our segmentation approach.

Table 6.5 presents classification results for *dynamic objects* subtree considering 3 categories: *pedestrians*, *2 wheeler* and *4+ wheeler*. Our overall accuracy is 99.34 %, which proves the good performance of our method. *4+ wheeler* such as cars are correctly classified with f_{mean} equal to 99.86 %. Note that *4+ wheeler* class contains 97.20 % of all 3D points in *dynamic object* node. For *pedestrians* and *2 wheeler* the f_{mean} are equal to 83.87 % and 75.71 %, respectively. The main drawback is that motorcycles may be over-segmented (as explained in Section 6.5.1) and then wrongly classified as pedestrians. Another problem is due to pedestrians walking too close to high cars (*e.g.* vans or small trucks), which may not be correctly separated leading to under-segmentation problems.

6.8.1.1 Comparison with the state of the art

A recent publication by Vallet et al. (2014) evaluates the current state of the art in urban scene analysis from MLS data. Results correspond to the TerraMobilita/iQmulus benchmark presented on July 8th, 2014 in Cardiff (UK), in conjunction with SGPâĂŹ14. For practical reasons, the benchmark only consisted in one of the ten zones of the TerraMobillita/iQmulus database, the "Cassette_idclass.ply" file.

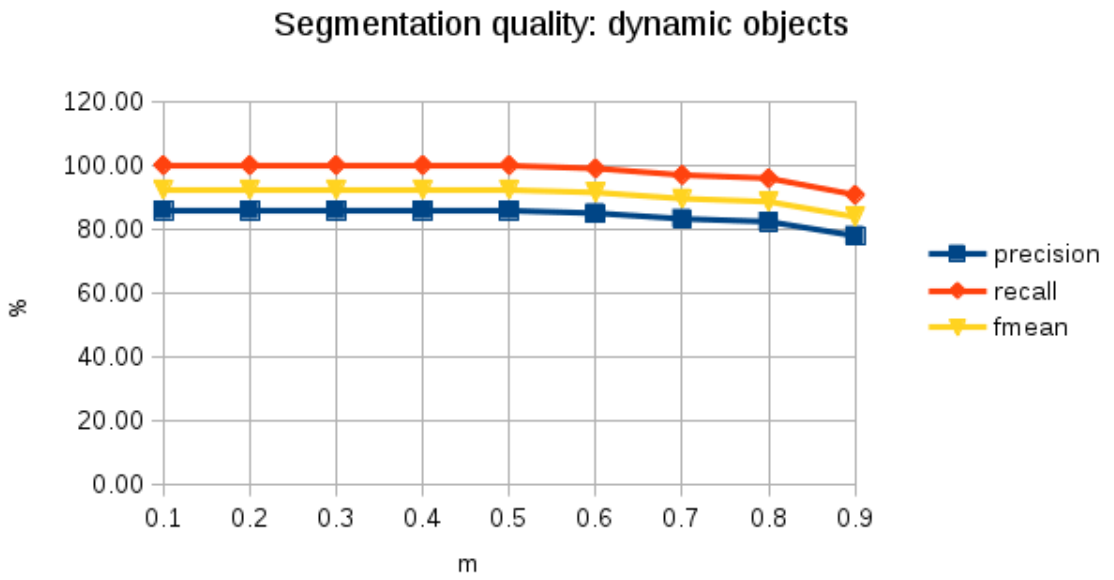


Figure 6.17: Segmentation quality for *dynamic object* class on the TerraMobilita/iQmulus database. The total number of *dynamic objects* annotated in the GT is 113. According to the results, our segmentation method correctly retrieves 97 dynamic objects (P(0.1)=85.84 %) for m=0.1, while 88 dynamic objects (P(0.9)=77.88 %) are retrieved for m=0.9. The geometric quality of our segmentation is good since the performance decays slowly. For example, $f_{\text{mean}}(m)$ decays from 92.38 % to 83.81 % for m varying from 0.1 to 0.9. Note that in the range m=[0.1, 0.5] our performances are constant, while in the range m=[0.5, 0.9] the $f_{\text{mean}}(m)$ decays less than 9%, which proves the robustness of our segmentation process.

Table 6.5: Classification results for *dynamic object* subtree on TerraMobilita/iQmulus database. GT: ground truth, AR: automatic result. In the confusion matrix, results are presented as percentages with respect to the total number of points in the 3D point cloud (12 million points).

GT/AR	**pedestrian**	**2 wheeler**	**4+ wheeler**	**Sum**	Recall	Precision	f_{mean}
pedestrian	1.63 %	0.00 %	0.12 %	1.76 %	92.80 %	76.51 %	83.87 %
2 wheeler	0.39 %	0.65 %	0.00 %	1.04 %	62.71 %	95.51 %	75.71 %
4+ wheeler	0.11 %	0.03 %	97.06 %	97.20 %	99.86 %	99.87 %	99.86 %
Sum	2.13 %	0.68 %	97.18 %	100.0 %	**Overall accuracy: 99.34 %**		

In the framework of that work, our method (called CMM method) is compared against that proposed by Weinmann et al. (2014) (called KIT method) and the authors discuss the benchmark results as follows:

> "It is quite obvious from the results that the CMM method outperforms the KIT method in all aspects. The explanation is quite simple: the KIT method is a point based classification only using a local information (neighborhood analysis) to make its decision. As many classes are composed of objects a neighborhood base method fails to classify all the objects points in the correct object class. This issue is discussed in Shapovalov et al. (2010). The classification performances are quite high, but we have to keep in mind that the numbers are computed only on the points that were classified in the same class of the higher level (for instance, for mobile vs static object, we only count the point classified as objects in the Ground Truth and the Algorithm Result)."

As aforementioned, the evaluation has been independently carried out by IGN. The reader is encouraged to review the publications by Brédif et al. (2014); Vallet et al. (2014) and to visit benchmark website: http://data.ign.fr/benchmarks/UrbanAnalysis/.

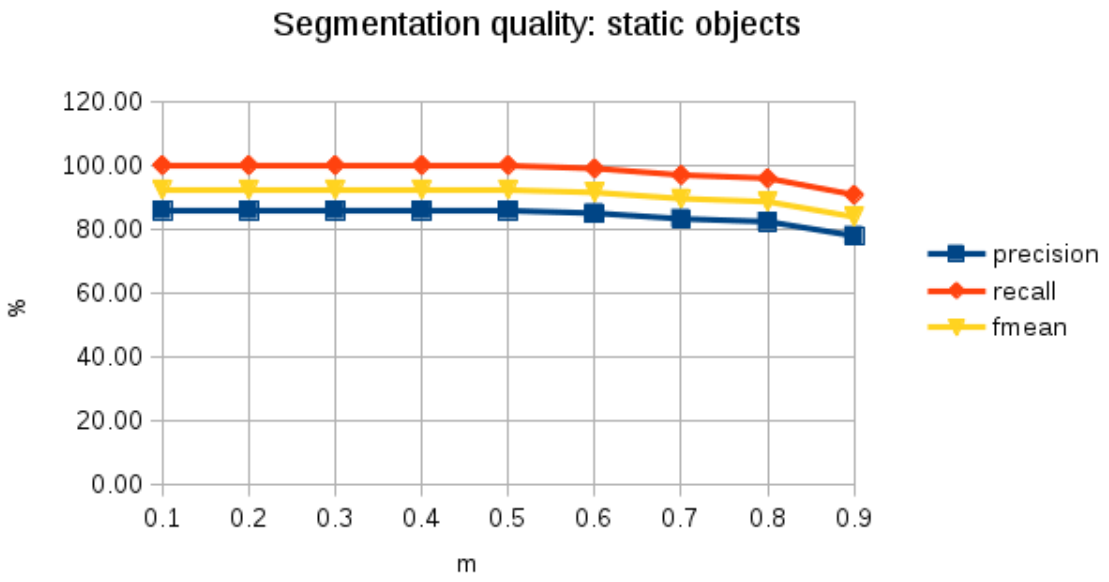


Figure 6.18: Segmentation quality for *static object* class on the TerraMobilita/iQmulus database. The geometric quality of our segmentation is good since the performance decays slowly. For example, $f_{\text{mean}}(m)$ decays from 87.80 % to 78.05 % for m varying from 0.1 to 0.9. Note that in the range m=[0.1, 0.5] our performances are constant, while in the range m=[0.5, 0.9] the $f_{\text{mean}}(m)$ decays less than 10%, proving the robustness of our segmentation approach.

6.8.2 Results: Paris-rue-Soufflot database

For this experiment, we use a manually annotated dataset from *rue Soufflot*, a street approximately 500 m long in the 5^{th} Parisian district. Acquisition was done by Stereopolis MLS system from IGN (Paparoditis et al., 2012), in the framework of TerraNumerica project (CapDigital, 2009). A typical scene is shown in Figure 6.19. It contains pedestrians, cars, lampposts, motorcycles, among others. This database was firstly used by Hernández and Marcotegui (2009c) to classify objects in four categories: cars, lampposts, pedestrians and other. However, their original annotation is no longer available. For the sake of comparison, we have manually annotated the database again[2] and managed to reproduce results consistent with those reported by the authors (shown in brackets in Table 6.6).

Data have been separated into two parts, training and test sets. This separation has been randomly done keeping 50% of the objects of each class in the training set and the rest in the test set. Color is not available in this database, thus only geometrical and contextual features have been used. In a first attempt, a single SVM classifier has been trained for all available categories. Training errors have been computed using 10-fold cross validation and high confusion rates were found between heterogeneous classes and classes with few examples, as shown in Figure 6.20(a). To overcome these problems, the hierarchical classification proposed in Section 6.6 is applied, as shown in Figure 6.20(b). The first SVM classifies well-discriminated objects (f_{mean}(train) greater than 80%), while the second one is exclusively dedicated to classes with higher confusion rates (f_{mean}(train) lower than 80%). Table 6.6 presents our classification results on the test set.

Our main contribution in the classification step is the use of contextual features and hierarchical SVM. With respect to Hernández and Marcotegui (2009c) work, classification results have been improved. On the one hand, cars and lamppost classification have the same maximal accuracy (100%) while the performance on the pedestrian class has been improved by about 15%. On the other hand, we use all available categories preserving the performance on cars and lampposts categories. The main problems appear with *furniture* and *other* classes because they are very heterogeneous. The same problem appears for *traffic lights* and *trash cans* classes because there are not enough samples in the database (4 and 5 samples, respectively).

[2] Paris-rue-Soufflot database is available at: http://cmm.ensmp.fr/~serna/downloads.html

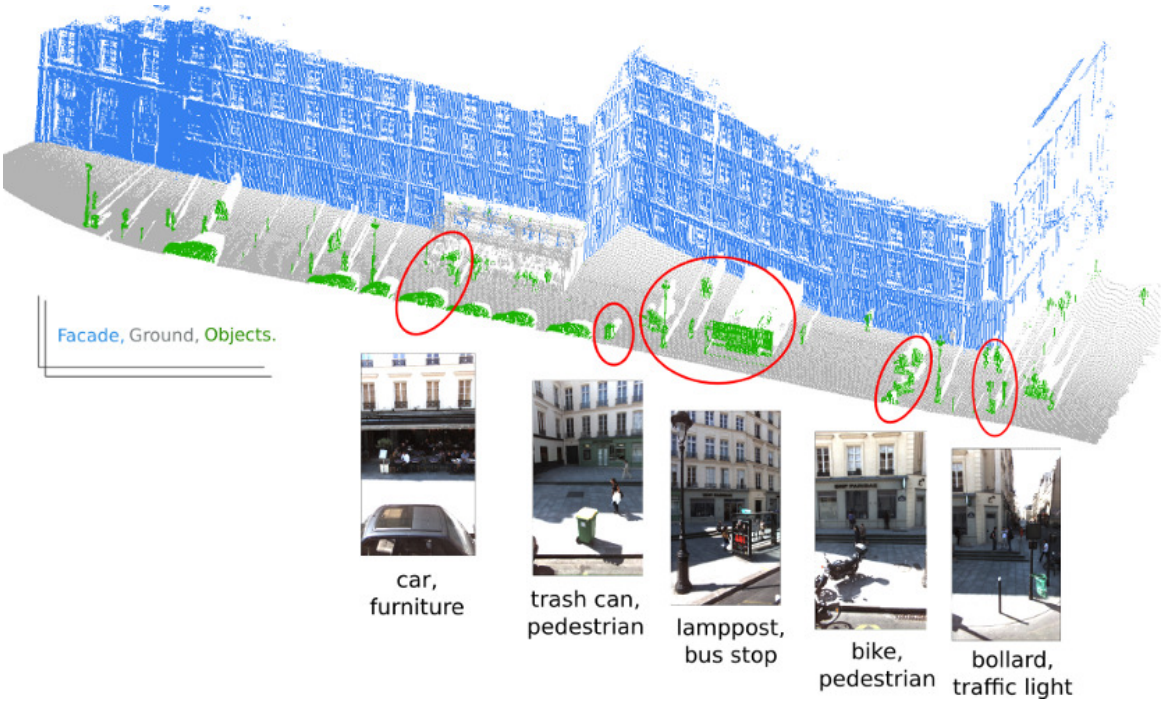


Figure 6.19: Example of urban objects manually annotated in Paris-rue-Soufflot dataset. It contains pedestrians, cars, lampposts, motorcycles, among others. For the sake of comparison, we have manually annotated the database and results are shown in Table 6.6. Acquired by Stereopolis system, IGN France.

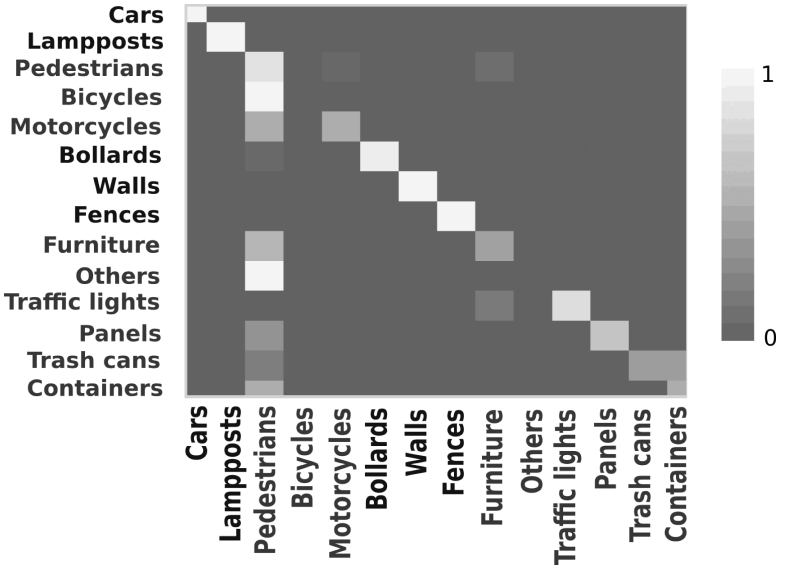
Table 6.6: Classification results on Paris-rue-Soufflot test set. In brackets results from Hernández and Marcotegui (2009c).

Class name	**Samples**	**Precision** (%)	**Recall** (%)	$\mathbf{f_{mean}}$(%)
Cars	27	100 (100)	100 (100)	100 (100)
Lampposts	12	100 (100)	100 (100)	100 (100)
Bollards	39	89	100	94
Walls	12	100	100	100
Fences	5	100	100	100
Pedestrians	101	86 (70)	84 (71)	85 (71)
Bikes	14	100	54	70
Furniture	30	67	67	67
Other	23	50	100	66.6
Traffic lights	4	0	0	0
Panels	7	100	100	100
Trash cans	5	0	0	0

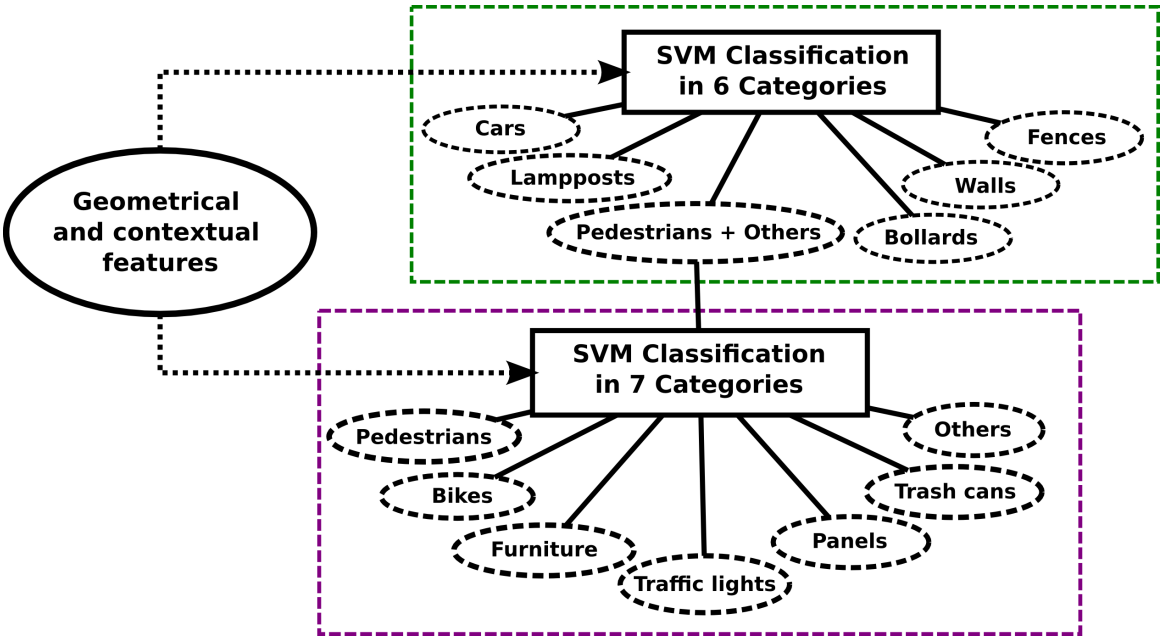
6.8.3 Results: Ohio database

Ohio database has also been used by Golovinskiy et al. (2009) and Velizhev et al. (2012) in order to evaluate their semantic analysis methods. This dataset is a combination of ALS and TLS data scanned in Ottawa city (Ohio, USA). It contains 26 tiles, 100 × 100 meters (approximately $4{\times}10^6$ points) each, as shown in Figure 6.21. A typical scene contains trees, cars, lampposts, among others. The GT consists in a labeled point marking the center of each object and its class.

Since our method is sequential, *i.e.* the input of each processing step is the output of the previous one, its evaluation is carried out in the same way. First, the detection process is applied to the entire database. Second, detected objects are used as input for the segmentation step. Third, correctly segmented objects are separated in two subsets (train and test) in order to perform the classification. Let us to explain each processing step and



(a) Confusion matrix using all available categories on the training set.



(b) Hierarchical classification

Figure 6.20: Hierarchical SVM classification on Paris-rue-Soufflot dataset. The first SVM classifies well-discriminated objects (f_{mean}(train) greater than 80%), while the second one is exclusively dedicated to classes with higher confusion rates (f_{mean}(train) lower than 80%).

its evaluation.

6.8.3.1 Evaluation: Detection

In order to evaluate our detection approach, an object is considered to be correctly detected if its GT center is included in the object hypotheses mask (Subsection Section 6.4), *i.e.* it has not been suppressed by any preprocessing filter and it has not been wrongly merged with the ground. Note that an object hypothesis may contain several connected objects or only a partial object. In the detection step, we are interested in keeping as much objects as possible, avoiding false alarms. This is important because non detected objects cannot be recovered in the subsequent steps. Table 6.7 presents the percentage of retrieved objects in this database. Our

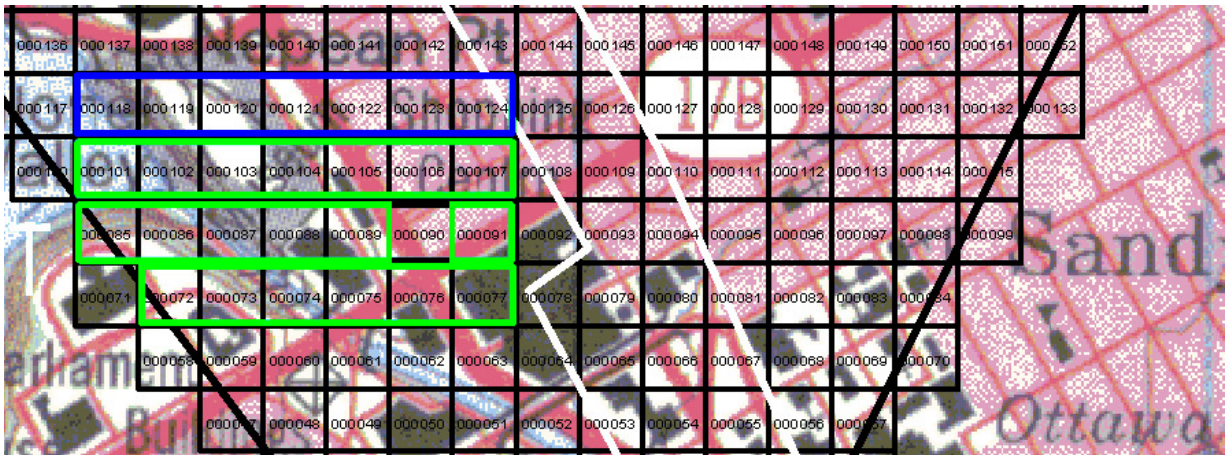


Figure 6.21: Ottawa city, Ohio (USA). The database contains 26 annotated tiles 100×100 meters each. A typical scene contains trees, cars, lampposts, among others. The GT consists in a labeled point marking the center of each object and its class. This database contains three types of tiles: training (blue), test (green) and non-annotated (black).

detection method retrieves 98% of objects, which outperforms other methods reported in the literature (92% by Golovinskiy et al. (2009) and 96% by Velizhev et al. (2012)). The number of false alarms cannot be estimated because many objects located on building roofs and in the forest are detected by our method (since they are real objects), but they have not been annotated in the database. Figure 6.22 shows the detection results on the 3D point cloud.

6.8.3.2 Evaluation: Segmentation

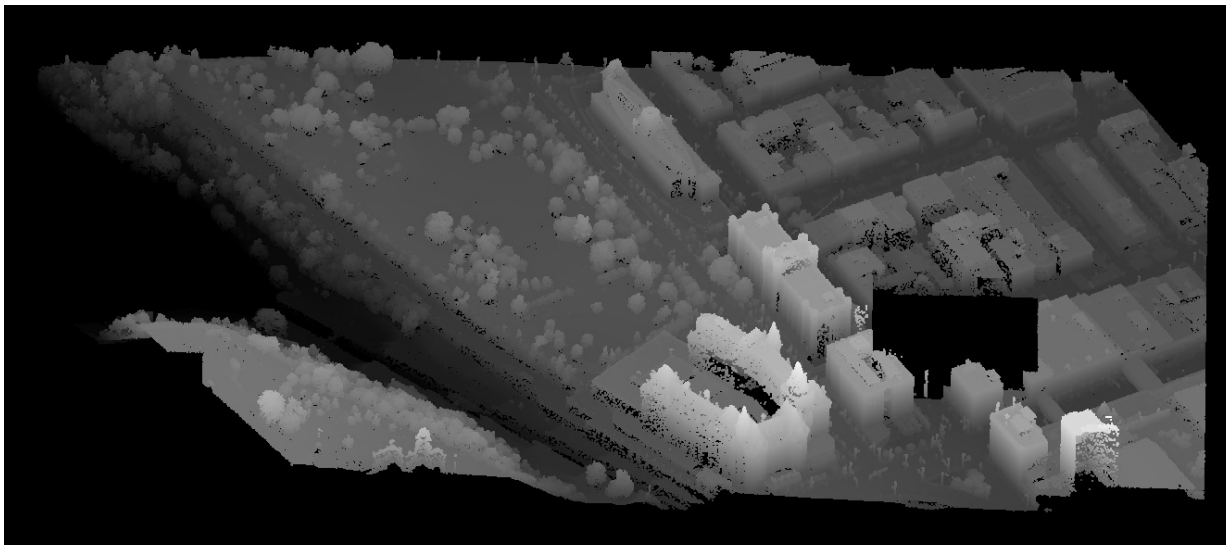
In order to evaluate our segmentation approach, an object is considered to be correctly segmented if it is isolated as a single object, *i.e.* connected objects are correctly separated (there is no under-segmentation) and each individual object is inside one and only one CC (there is no over-segmentation). However, an estimation of under-segmentation and over-segmentation errors cannot be done on Ohio database because it only contains a GT point for each object. In that sense, an object is considered to be correctly segmented if it is marked with one and only one GT point.

As shown in Table 6.7, our method segments correctly 76% of detected objects. Objects such as cars, lampposts, parking meters and signs are correctly segmented (Recall greater than 80%). The main problem comes from under-segmentation of connected objects such as light poles, posts and trees. Since this kind of clusters has only one maximum on the elevation image (the highest object), they are not correctly segmented by our method. Note that trees recall is 90%, which is considered as a satisfactory segmentation. However, trees represent approximately 34% of the objects in the database, which implies that under-segmented trees affect seriously the recall of other classes, in particular for classes with few objects.

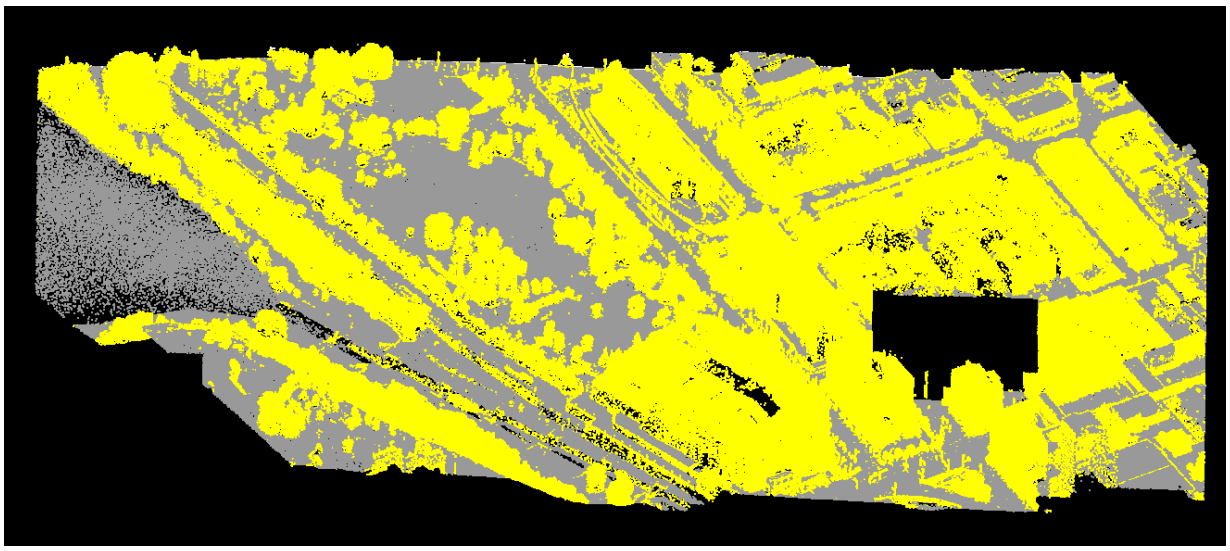
6.8.3.3 Evaluation: Classification

For the classification experiments, segmented objects in the north quarter of the city (7 tiles, 458 objects) are used for training and the rest (19 tiles, 677 objects) for testing. Training and testing tiles are the same as in (Golovinskiy et al., 2009), for comparison purposes. The number of objects per class on both training and test sets are detailed in Table 6.9.

Geometrical, contextual and color features (Subsection Section 6.6) are combined in this experiment in order to define the best classification features. Classification performance obtained using different combinations of them is given in Table 6.8. The best overall accuracy (82%), defined as the ratio between the number of correctly classified objects and the total number of objects, is obtained combining geometrical and contextual features. Detailed results are presented in Table 6.9. It is noteworthy that including color information degrades the classification accuracy. The reason is that in this database, color information is the result of overlapping ALS and TLS. During acquisitions, color sensors were not calibrated, thus their superposition is not perceptually coherent, as shown in Figure 6.23.



(a) Z-coordinate



(b) Object detection and DTM generation

Figure 6.22: Ohio database: object detection (yellow) and DTM generation (gray). In order to evaluate our detection approach, an object is considered to be correctly detected if its GT center is included in the object hypotheses mask, *i.e.* it has not been suppressed by any preprocessing filter and it has not been wrongly merged with the ground. Note that an object hypothesis may contain several connected objects or only a partial object. In the detection step, we are interested in keeping as much objects as possible, avoiding false alarms. This is important because non detected objects cannot be recovered in the subsequent steps.

Table 6.9 presents detailed classification results. Precision, Recall and f_{mean} for each class are presented. In this experiment, classes with less than 5 objects, either in the training set or in the testing set, are not considered in the classification process. Therefore, only 6 categories have been used. It is noteworthy that cars, trees and posts are correctly classified. However, lampposts, lights and signs classification has lower performance.

For the sake of clarity, Table 6.10 shows the confusion matrix. Note that cars are correctly classified while lampposts, lights, posts, signs and trees are mixed up, which is comprehensible because they are pole-like objects.

In an attempt to solve confusion problems, the hierarchical classification approach (proposed in Section 6.6) has been studied. Lampposts, lights, posts and signs have been put together in a new class, while cars and trees

Table 6.7: Detection and segmentation results on Ohio dataset.

			Detection		Segmentation	
Class	**Name**	**GT samples**	**Detected**	**Recall**	**Segmented**	**Recall**
1	Ad cylinder	6	6	100 %	5	83 %
2	Bush	29	28	97 %	23	82 %
3	Car	240	237	99 %	195	82 %
4	Dumpster	1	1	100 %	1	100 %
5	Fire hydrant	19	16	84 %	13	81 %
6	Flagpole	2	2	100 %	2	100 %
7	Lamppost	146	143	98 %	117	82 %
8	Light pole	62	60	97 %	46	77 %
9	Mailing box	4	4	100 %	1	25 %
10	Newspaper box	42	35	83 %	5	14 %
11	Parking meter	10	10	100 %	10	100 %
12	Post	377	376	100 %	208	55 %
13	Recycle bin	6	6	100 %	3	50 %
14	Sign	96	92	96 %	79	86 %
15	Telephone booth	4	4	100 %	2	50 %
16	Traffic control box	8	5	63 %	2	40 %
17	Traffic light	42	42	100 %	34	81 %
18	Trash can	19	19	100 %	8	42 %
19	Tree	552	543	98 %	490	90 %
20	Box transformer	2	2	100 %	0	0 %
	Total	**1667**	**1631**	**98 %**	**1244**	**76 %**

Table 6.8: Classification accuracy on Ohio dataset using different features combination.

Features	**Overall accuracy**
Geometrical	75%
Geometrical + C_{ind}	77%
Geometrical + C_{ind} + N_{neigh}	**82%**
Geometrical + C_{ind} + N_{neigh} + Color	72%

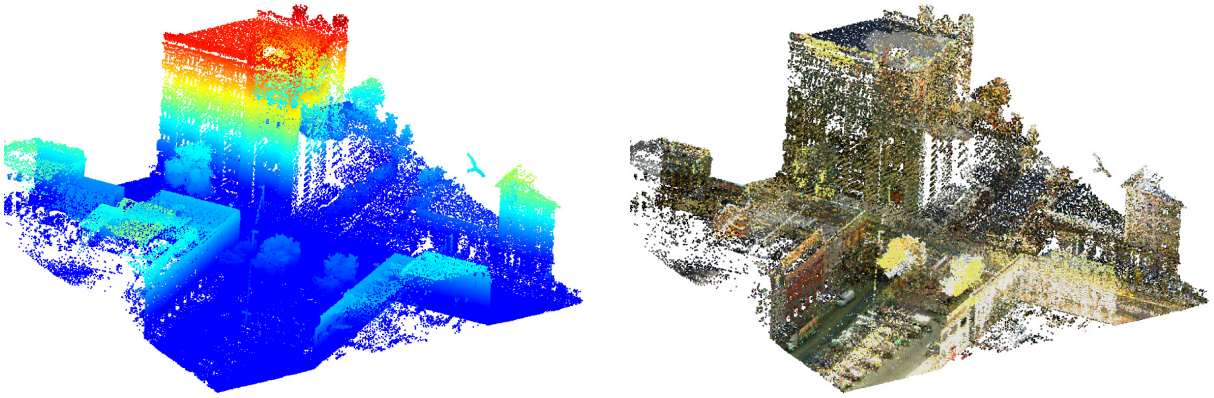
are preserved in their original classes. A first classifier is applied to separate correctly discriminated objects, and a second one is exclusively dedicated to classes with higher confusion rates. After our experiments, we have noted that this approach does not provide any global improvement in this database since f_{mean} increases by 16% for lampposts and lights, but it decreases by 15% for posts and signs. The conclusion here is that a hierarchical approach is not enough to solve confusion problems since objects are too similar. A possible solution is the use of other features which allow a clearer separation between classes.

Table 6.11 presents results gathering lampposts, lights, posts, and signs in a more general category called pole-like objects. With 3 classes, the overall accuracy rises up to 88%.

6.8.3.4 Comparison with the state of the art

Ohio database has been chosen because it contains many different objects, it is large enough to exemplify a large-scale application, and comparison with the state of the art is possible since it has been used in other works (Golovinskiy et al., 2009; Velizhev et al., 2012).

We present our results on 26 tiles. However, in the original publication by Golovinskiy et al. (2009) (the website containing the dataset is not longer available), they report 27 tiles. Therefore, the number of objects is not the same due to this missing tile. Additionally, some important differences have been noticed with respect to the aforementioned authors: on the one hand, with respect to Velizhev et al. (2012), they have only used 2 classes (cars and light poles), thus only a partial comparison can be done; on the other hand, with respect to



(a) Tile #72: Height information

(b) Tile #72: Color information

Figure 6.23: Color and height information on Ohio database. It is noteworthy that including color information degrades the classification accuracy. The reason is that in this database, color information is the result of overlapping ALS and TLS. During acquisitions, color sensors were not calibrated, thus their superposition is not perceptually coherent.

Table 6.9: Classification results on Ohio database. Classes with less than 5 objects, either in the training set or in the testing set, are not considered. Pred: predicted, TP: true positives, P: Precision, R: Recall.

		Correctly segmented		**Classification**				
Class	**Name**	**Train**	**Test**	**Pred**	**TP**	**P**	**R**	$\mathbf{f_{mean}}$
1	Ad cylinder	2	3					
2	bush	1	22					
3	car	108	87	85	75	88%	86%	87%
4	dumpster	1	0					
5	Fire hydrant	3	10					
6	flagpole	1	1					
7	Lamppost	33	84	78	51	65%	61%	63%
8	Light pole	14	32	22	16	73%	50%	59%
9	Mailing box	0	1					
10	Newspaper box	0	5					
11	Parking meter	10	0					
12	post	132	76	85	66	78%	87%	82%
13	Recycle bin	1	2					
14	sign	34	45	44	33	75%	73%	74%
15	Telephone booth	1	1					
16	Traffic control box	1	1					
17	Traffic light	4	30					
18	Trash can	0	8					
19	tree	137	353	363	317	87%	90%	89%
20	Box transformer	0	0					
Total (used classes)		**458**	**677**	**677**	**558**	**82%**	**82%**	**82%**
Total (all objects)		**483**	**761**					

Golovinskiy et al. (2009), the main difference comes from the fact that they do not consider trees nor bushes in their analysis.

Table 6.12 presents a quantitative comparison with the state of the art. Taking into account only 6 categories, the ones used during classification. Our detection method (accuracy equal to 99%) performs better than the

Table 6.10: Confusion matrix for classification in 6 classes on Ohio dataset.

GT\Predict.	Cars	Lampposts	Light	Post	Sign	Tree	**Total**
Car	**75**	0	0	0	1	11	87
Lamppost	1	**51**	1	11	1	19	84
Light	0	6	**16**	0	0	10	32
Post	0	3	1	**66**	2	4	76
Sign	0	3	0	7	**33**	2	45
Tree	9	15	4	1	7	**317**	353
Total	85	78	22	85	44	363	

Table 6.11: Confusion matrix gathering lampposts, lights, posts, and signs in the same category. Results on Ohio dataset.

GT\Predict.	Cars	Pole-like	Trees	Total	**Precision**	**Recall**	$\mathbf{f_{mean}}$
Car	**75**	1	11	87	88%	86%	87%
Pole-like	1	**201**	35	237	88%	85%	86%
Trees	9	27	**317**	353	87%	90%	89%
Total	85	229	363				

other two reported in the literature. Our classification accuracy is equal to 82%, whereas Golovinskiy et al. (2009) correctly classify 65% of the objects considered by their method. With respect to the segmentation method, results from Velizhev are not available and our accuracy (78%) is 8% lower than that reported by Golovinskiy et al. (2009). On the one hand, our major under-segmentation problem is due to clusters formed by trees and pole-like objects, where the highest object is the only significant maximum. On the other hand, our major over-segmentation problem is when segmenting objects with several regional maxima such as trees. To summarize, our sequential method correctly detects, segments and classifies 99%×78%×82% = 64% of the annotated objects.

Table 6.12: Summarized comparison with other methods reported in the literature on Ohio dataset. Percent values indicate the accuracy in each step of the semantic analysis.

	Golovinskiy et al. (2009)	Velizhev et al. (2012)	Serna and Marcotegui (2014)
Detection	92%	96%	99%
Segmentation	86%	N/A	78%
Classification	65%	67%	82%
Overall accuracy			**64%**
Computational time	7.3 min/tile (3 GHz PC)	5 ∼ 10 min/tile (4×2.4 GHz PC)	1 min/tile (4×2.4 GHz PC)

With respect to computational time (last row in Table 6.12), our method is up to 10 times faster than the other two works. These three works use general-purpose machines and they are not specially optimized nor parallelized. The aim of this comparison is to give an idea to the reader about the computational time and the potential for large-scale or other time-constrained applications. One of the reasons of our faster processing is due to the use of elevation images and image processing algorithms since their computational cost is less expensive than direct 3D processing. Note that the typical speed of a MLS system is 30 km/h, which corresponds approximately to a covered area of 10,000 m^2/minute on a 20 m wide street without considering stops nor traffic lights. In this database, our processing speed is 10,000 m^2/minute. This is a very fast off-line processing since acquisition and processing times are equal.

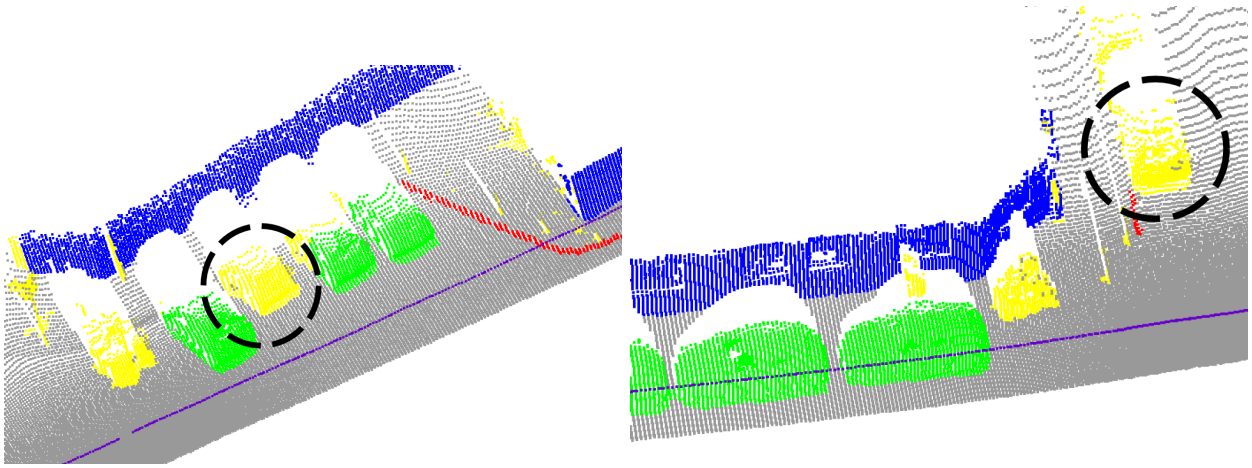
6.8.4 Results: Paris-rues-Vaugirard-Madame database

Dealing with cars has a particular interest in the framework of TerraMobilita project since one of the applications consists in computing automatic parking statistics, as presented in Section 1.3. In order to evaluate the potential of an automatic method, several 3D point clouds of the same street section in Paris (rues Vaugirard and rue Madame, approximately a 500 m long section) have been acquired at different hours. Acquisition was done by Stereopolis MLS system from IGN (Paparoditis et al., 2012). Then, we apply our automatic methodology in order to detect, segment and classify cars. For the classification evaluation, urban objects were manually annotated. We use 2307 objects (129 *cars* and 2178 *other*) as training set, and 970 objects (53 *cars* and 917 *other*) as testing set. Note that a hierarchical classification is not applied since we are only interested in cars.

Color information is not available. Therefore, only geometrical and contextual features have been used. Table 6.13 presents our classification results using a binary SVM. The performance of our method is proved since 99.7% of the objects are correctly classified. Note that 5.4% of the cars have not been properly identified due to occlusion problems, as shown in Figure 6.24. Note that these cars are perpendicularly parked with respect to the acquisition trajectory, therefore only a part of them has been scanned. Note that this database contains a few number of mobile cars because only one laser scanner oriented to the right sidewalk has been used.

Table 6.13: Car classification results on Paris-rues-Vaugirard-Madame database.

Class name	Precision	Recall	f_{mean}
Cars	100.0 %	94.6 %	97.2 %
Other	99.7 %	100.0 %	99.9 %

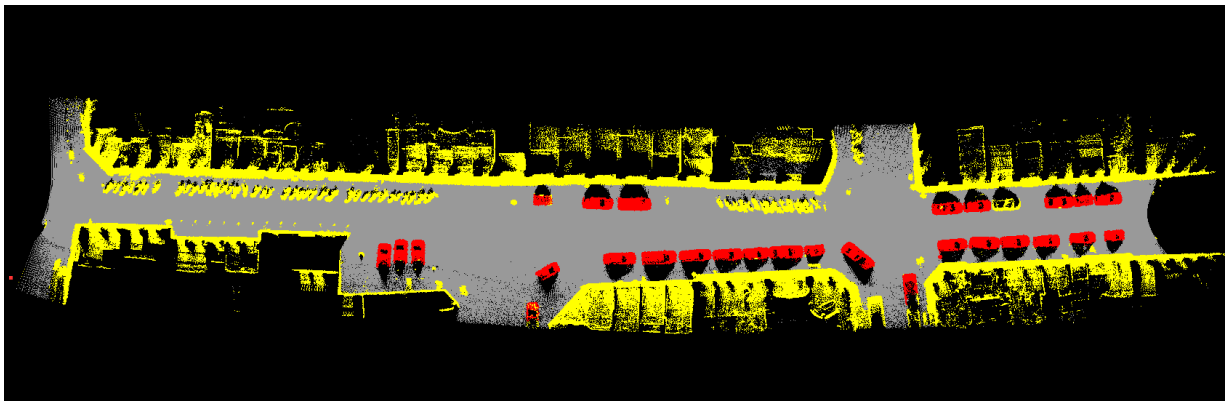


(a) (b)

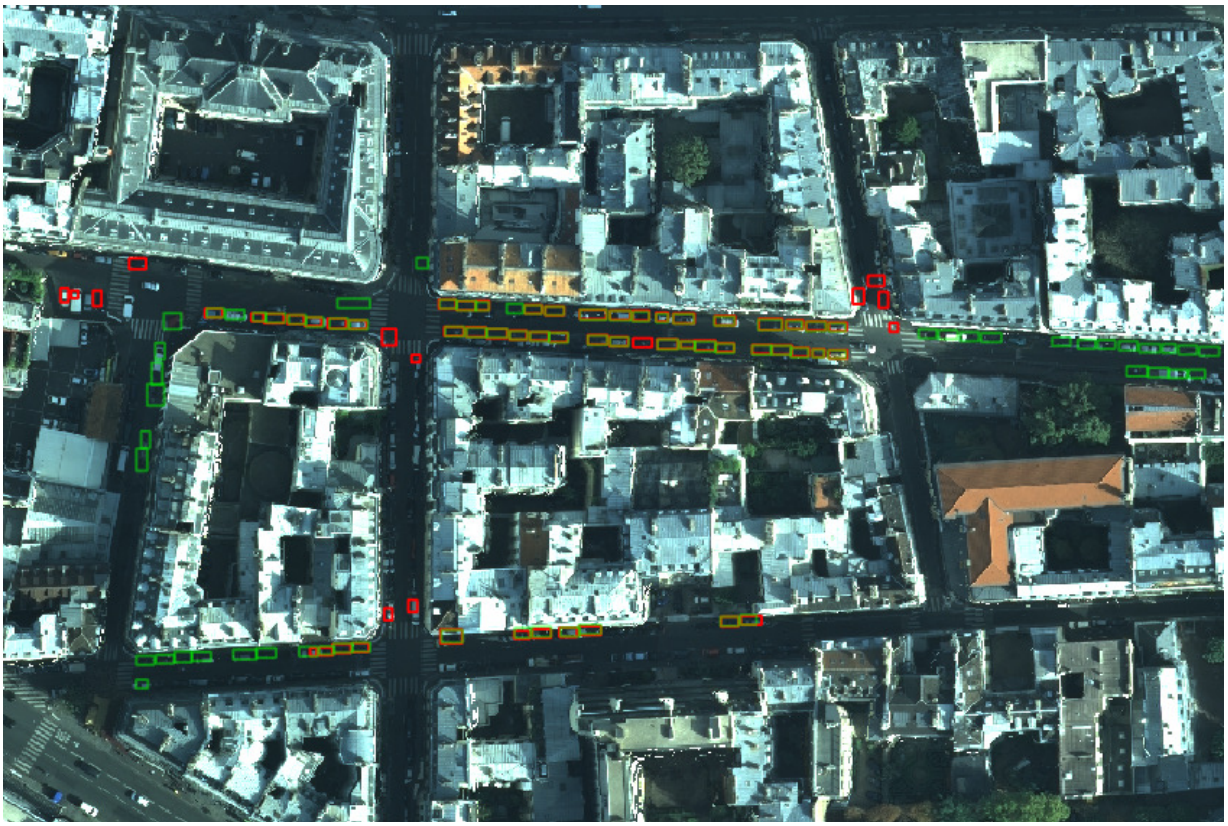
Figure 6.24: Classification errors due to occluded cars parked perpendicularly to the acquisition trajectory, therefore only a part of them has been scanned. Facade (blue), ground (gray), cars (green), trajectory (magenta), curbs (red), other (yellow). Point clouds taken from Paris-rues-Vaugirard-Madame database.

In order to demonstrate that our method can be easily generalized, we have used a classifier trained on Ohio dataset in order to classify cars on Paris-rues-Vaugirard-Madame database. A f_{mean} equal to 90.0% has been obtained. This result is slightly lower than that reported in Table 6.13 (97.2%). However, the great advantage is that a new annotation may not be required when working with a new database.

Other experiments have been carried out using 3D data acquired by L3D2 system by MINES ParisTech (presented in Section 2.4.2) on the same test zone in Paris. Figure 6.25(a) shows classification result reprojected onto the 3D point cloud. Figure 6.25(b) presents results exported as shapefiles to a GIS in order to visualize and compute parking statistics in an easier way.



(a) Classification result reprojected onto the 3D point cloud: cars (red), ground (gray), other (yellow). This point cloud corresponds to a segment of *rue Madame* in Paris, France. L3D2, CAOR-MINES ParisTech©.



(b) Classification results exported to a GIS. Cars detected in the first passing (red), in the second (green), in both (orange).

Figure 6.25: Classification results on a 3D data acquired by L3D2 system by MINES ParisTech in the *rue Madame* in Paris. (a) shows classification result reprojected onto the 3D point cloud. (b) presents results exported as shapefiles to a GIS in order to visualize and compute parking statistics in an easier way.

At this point, our system is able to correctly extract cars and compute additional information such as geometric features, geographic position and GPS time at the acquisition moment. However, a comparison between cars parked in the same place at different moments is required to compute parking duration statistics. To avoid confusions between those cars, geometrical and color features can be used. Additionally, relative sensor precision between different acquisitions becomes a critical issue. In efficiency terms, an automatic method seems to be suitable for this problem since the acquisition vehicle can go up to 20 times faster than a person. Additionally, the automatic processing takes only a few minutes and it is comparable to the acquisition time.

6.9 Conclusions

We propose an automatic and robust approach to detect, segment and classify urban objects from 3D point clouds. Processing is carried out using elevation images and the final result can be exported to a GIS and reprojected onto the 3D point cloud for displaying and post-modeling purposes.

One of the main drawbacks processing 3D urban data using elevation images is that high objects may occlude lower objects located below them. That is why we propose a segmentation strategy using two slices. In the lower slice, objects are detected using a two-fold strategy considering both structures connected to the boundary of the scene as well as ground discontinuities. Then, a filtering step is performed in order to reduce noise but preserving thin vertical structures. Subsequent, connected objects are segmented assuming that the number of significant maxima is equal to the number of connected objects. In the upper slice, a rule-based method has been proposed in order to segment facades, trees, poles and off-ground objects. Results from both slices are integrated based on connectivity on the slices boundary. Additionally, 3D connectivity and adaptive voxelization can be used as well.

It is obvious that processing two slices is more expensive than processing only one elevation image. Therefore, two slices are only used in databases containing several trees or other high objects occluding objects below them. In particular, this strategy has been successfully applied in the TerraMobilita/iQmulus database. Other databases such as Paris-rues-Vaugirard-Madame and Paris-rue-Soufflot do not contain trees in the public space, thus processing by slices has not been required. In the case of Ohio database, most trees correspond to a wood in the east side of the city, which has been acquired by ALS. Therefore, lower tree parts are not visible and processing by slices is not justified.

After segmentation, objects are classified in several categories using an SVM approach with geometrical and contextual features. Our geometrical features can be adapted to any XYZ point cloud. Thus, classification can be easily generalized, *i.e.* training on a database and testing on another one, as shown in Paris-rues-Vaugirard-Madame database. This is a significant advantage because the model learned from a database can be applied to another one, even acquired by a different acquisition system, without the tedious manual annotation.

In the case of TerraMobilita/iQmulus dataset, we have proposed a protocol in order to evaluate classification, detection and segmentation quality. Additionally, benchmark results have proved that our method is accurate and overcomes other works reported in the literature using the same database. Our results on Ohio dataset show that our method retrieves 99% of the objects in the detection step, 78% of connected objects are correctly segmented, and 82% of correctly segmented ones are correctly classified using geometrical and contextual features. On Paris-rue-Soufflot dataset, our proposed hierarchical classification leads to an improvement of about 15% on *pedestrian* class with respect to previous works while preserving a good performance in other classes. Moreover, new classes (not considered in previous works) have been taken into account. Other experiments have been also carried out on Paris-rues-Vaugirard-Madame datasets in order to exploit our classification results to compute automatic parking statistics.

Our method is proven to be robust to noise since small and isolated structures are eliminated using morphological filters. Additionally, it is fast because we project 3D points onto elevation images and we process them as a complete set using digital image processing techniques.

Even if our method presents good results and outperforms other state of the art methods, it is noteworthy that several improvements should be done before developing a mature application. Our main problem, common to all methods in the literature, is due to large occluded regions. Several scans of the same zone could reduce this problem. Some under-segmentation and over-segmentation problems have been also pointed out. A possible solution can include shape/texture analysis to help deciding whether an object should be re-segmented. Up to now, we have only used the spatial information available in the point cloud. However, additional features such as laser intensity and texture could improve our performance. Additionally, in the future we are planning to use Velodyne data in order to distinguish static from mobile obstacles and to reduce occlusion problems.

The TerraMobilita/iQmulus benchmark 2014 is still open, thus other authors can submit their results in order to get comparisons with the state of the art. As aforementioned, evaluation is independently carried out by IGN[3].

[3] If you are interested in participate in the benchmark, please contact Dr. Bruno Vallet for details uploading results: http://data.ign.fr/benchmarks/UrbanAnalysis/#Contact [Last accessed: July 23, 2014.]

7 Attribute-based filtering and segmentation

7.1 Résumé

Dans ce chapitre, nous présenterons des contributions à la morphologie mathématique dans le domaine des opérateurs basés sur des attributs. Nous montrerons certaines de leurs applications telles que la reconstruction, la morphologie adaptative, l'extraction de caractéristiques, le filtrage et la segmentation. Dans un premier temps, nous rappellerons des concepts basiques de la morphologie mathématique. Ensuite, nous exposerons une méthode de propagation contrôlée ainsi qu'une méthode de segmentation basée sur l'évolution d'attributs. Finalement, nous illustrerons par un exemple de segmentation des cellules allongées dans le cadre d'une application industrielle.

7.2 Introduction

Local operators constitute powerful techniques in digital image processing. They are based on the neighborhood of each pixel, defined by a kernel. In general, such neighborhood is defined by a ball of radius r centered at the point to be processed. In the digital case, the kernel is reduced to the definition of a local neighborhood describing the connections between adjacent pixels. In Mathematical Morphology (MM), these kernels are called structuring elements (SE) and they are the basis of sophisticated nonlinear techniques for filtering, feature extraction, detection and segmentation (Matheron, 1975; Serra, 1982, 1988, 1993). It has been shown that adaptive approaches can lead to important improvements (Lerallut et al., 2007; Maragos and Vachier, 2009; Pinoli and Debayle, 2009; Roerdink, 2009; Angulo, 2011).

In this chapter, several methodological contributions to mathematical morphology are presented. We have developed powerful attribute-based operators useful in a wide range of applications such as: attribute controlled reconstruction (Section 7.4), adaptive mathematical morphology (Section 7.5), feature extraction (Section 7.6), filtering and segmentation (Section 7.7). Besides, Chapter 5 presents an application to the semantic urban analysis on the segmentation of elongated facades. Several contributions of this chapter have already been published in Serna and Marcotegui (2013a); Serna et al. (2014a). Further details are given in the following sections.

This chapter is organized as follows. Section 7.3 revisits some basic concepts in MM: quasi-flat zones, threshold decomposition and attribute profiles. Section 7.4 defines our propagation controlled by the evolution of attributes. Section 7.5 presents an application to adaptive MM using input-adaptive SE. Section 7.6 proposes a feature extraction technique where input-adaptive SE are used to assess shape features on the image. Section 7.7 introduces a segmentation methodology based on the profile of a new attribute: the area-stable elongation. This methodology is successfully applied to segment elongated cells in fluorescence multiphoton microscopy images of engineered skin. Finally, Section 7.8 concludes the chapter.

7.3 Background

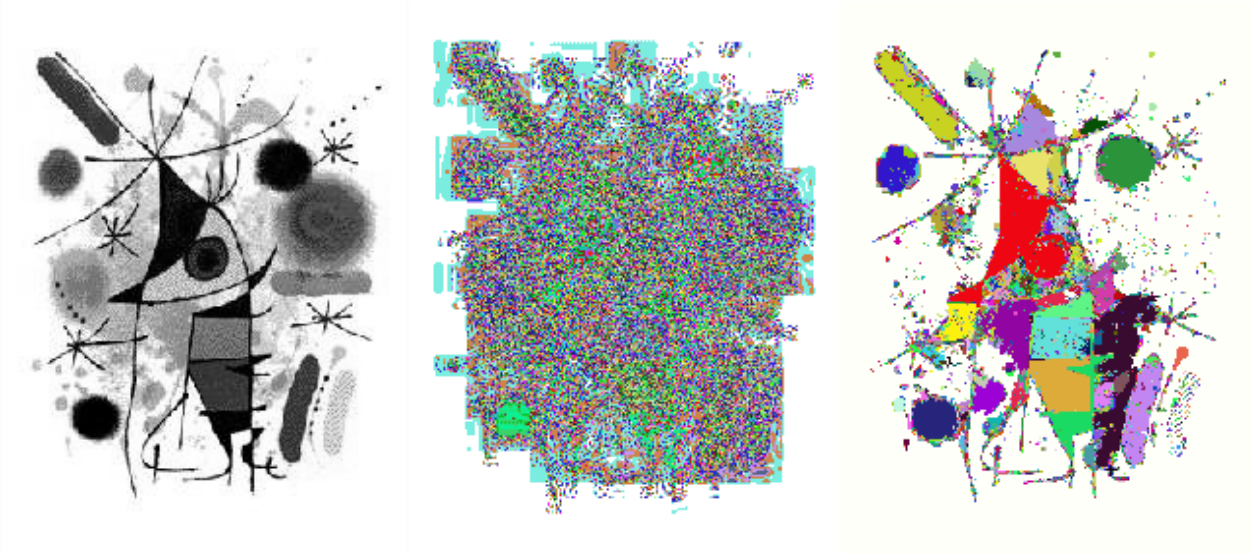
7.3.1 Quasi-flat zones

Connectivity relations naturally lead to partitions (Serra, 1998). For example, the connectivity relation induced by the equality of gray-level divides the image into maximal connected components (CC) of constant gray-level, called flat-zones (Salembier and Serra, 1995). In most cases, partition in flat zones results in too many segments, as shown in the example of Figure 7.1(b). A less restrictive connectivity relation can be defined adding a threshold λ. It allows to connect adjacent pixels if their gray-level difference does not exceed λ. This procedure, firstly introduced in image processing by Nagao et al. (1979), is called quasi-flat (or λ-flat) zones labeling and it is defined by Meyer (1998) as:

Definition 7.3.1 *Let I be a digital gray-scale image $I : D \to V$, with $D \subset Z^2$ the image domain and $V = [0, ..., R]$ the set of gray levels. Two neighboring pixels p, q belong to the same λ-flat zone of I, if their difference*

$|I(p) - I(q)|$ is smaller than or equal to a given λ value.

The definition of λ-flat zones is very useful in image partition, simplification and segmentation. An example of flat zones and quasi-flat zones labeling is shown in Figure 7.1.



(a) Original image (b) Flat zones (λ=0) (c) Quasi-flat zones (λ=15)

Figure 7.1: Example of flat and quasi-flat zones on a gray level image. Each color represents a segment in the partition. This is very useful for image partition, simplification and segmentation.

7.3.2 Threshold decomposition and attribute profile

A gray-scale image can be represented as a stack of binary images using threshold decomposition (Wendt et al., 1986; Maragos and Ziff, 1990) as defined in Definition 7.3.2:

Definition 7.3.2 *Let I be a digital gray-scale image $I : D \rightarrow V$, with $D \subset Z^2$ the image domain and $V = [0, ..., R]$ the set of gray levels. A decomposition of I can be obtained considering successive thresholds:*

$$T_t(I) = \{p \in D | I(p) > t\} \; \forall t = [0, ..., R-1] \tag{7.1}$$

Since this decomposition satisfies the inclusion property $T_t(I) \subseteq T_{t-1}(I), \forall t \in [1, ..., R-1]$, it is possible to build a tree, called the component tree, with level sets $T_t(I)$. Each branch of the tree represents the evolution of a single connected component X_t. An attribute profile is the evolution of an attribute (*e.g.* area, perimeter, elongation, average gray-level, etc.) of a given CC along a branch of the tree.

Figure 7.2 illustrates the threshold decomposition for a 1D function, its component tree and the attribute (width) profiles for the two function maxima (p_A and p_B). Events on this attribute profile are useful to segment objects (Jones, 1999), extract features (Pesaresi and Benediktsson, 2001; Beucher, 2007; Morard et al., 2011b) and define adaptive structuring elements (Serna and Marcotegui, 2013a).

7.3.3 Attributes: Geodesic elongation

In general, there are two types of attributes: increasing and non-increasing (Breen and Jones, 1996). On the one hand, an attribute is *increasing* when its value is greater or equal to the attribute computed on any subset of the object, *i.e.* an increasing attribute computed on a node of the component tree is greater than or equal to the attribute computed on any child of the same node. The most common increasing attribute is the area, used to compute area openings (Vincent, 1994) and area stability (Matas et al., 2004). On the other hand,

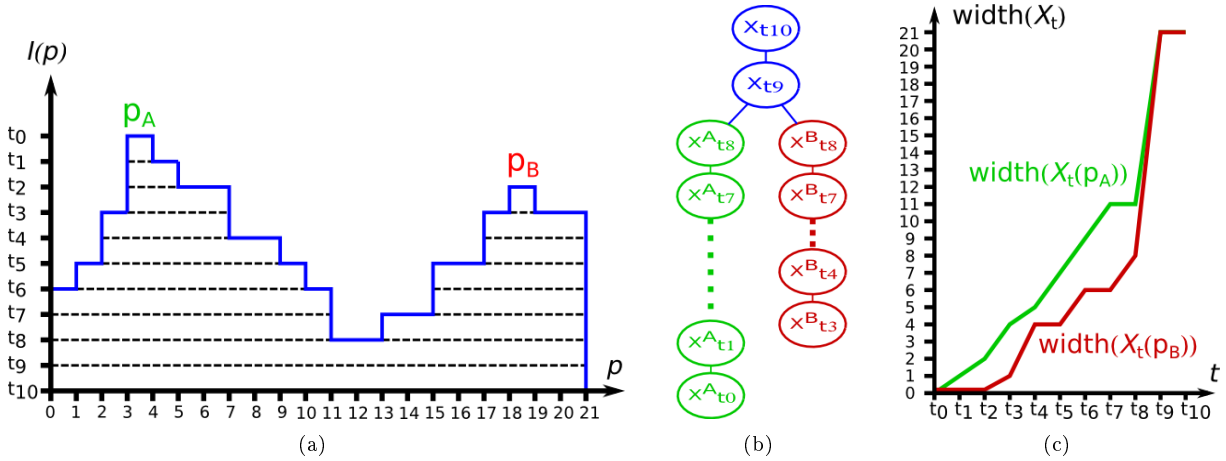


Figure 7.2: (a) 1D threshold decomposition; (b) component tree; (c) attribute profile for a 1D function from its two maxima (p_A and p_B). Events on this attribute profile are useful to segment objects (Jones, 1999), extract features (Pesaresi and Benediktsson, 2001; Beucher, 2007; Morard et al., 2011b) and define adaptive structuring elements (Serna and Marcotegui, 2013a).

an attribute is *non-increasing* when the latter property does not hold. In general, shape attributes –such as circularity, tortuousity, elongation, among others– are non-increasing and scale-invariant.

In this thesis, we focus on geodesic elongation (Lantuéjoul and Maisonneuve, 1984), simply called henceforth elongation. The elongation $E(X_t)$ of an object X_t is a shape descriptor useful to characterize long and thin structures. It is proportional to the ratio between square geodesic diameter $L^2(X_t)$ and object area $S(X_t)$, as shown in Equation (7.2). The geodesic diameter $L(X_t) = \sup_{x \in X_t}\{l_x(X_t)\}$ is the length of the longest geodesic arc of X_t, *i.e.* the longest internal segment $l_x(X_t)$ connecting the two end points of X_t (Lantuéjoul and Beucher, 1981). Figure 7.3 illustrates the definition of the geodesic diameter.

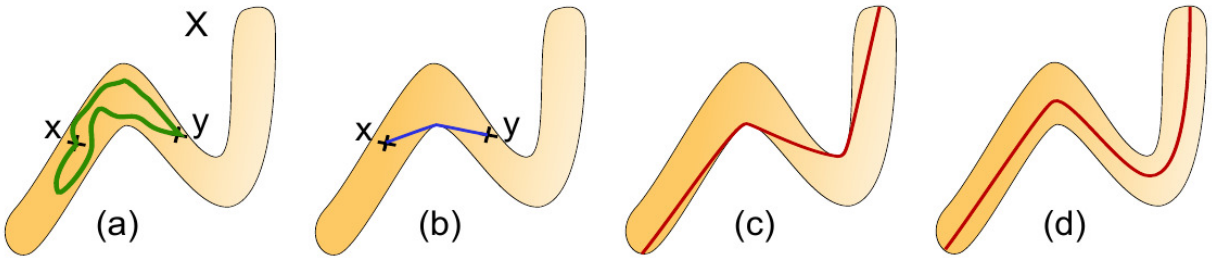


Figure 7.3: Geodesic diameter $L(X)$ definition. (a) Two paths between points x and y; (b) geodesic arc between these two points; (c) longest geodesic arc of object X, whose length is the geodesic diameter $L(X)$; (d) generalized geodesic distance, longest geodesic arc. Image taken from Morard et al. (2013)

$$E(X_t) = \frac{\pi}{4}\frac{L^2(X_t)}{S(X_t)} \tag{7.2}$$

The longer and narrower the object, the higher the elongation. The lowest bound is reached with the disk, where $E(\text{disk}) = 1$. An example of elongation for binary objects is presented in Figure 7.4. The number on each object corresponds to its approximated elongation. An efficient implementation can be found in Morard et al. (2013).

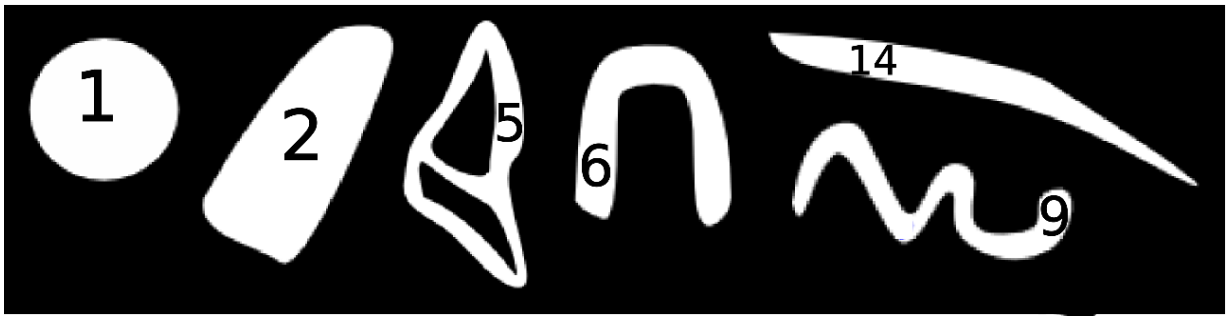


Figure 7.4: Geodesic elongation for different binary objects. The elongation values have been approximated to get integer values. The longer and narrower the object, the higher the elongation. The lowest bound is reached with the disk, where $E(\text{disk}) = 1$. In this image, the number on each object corresponds to its approximated elongation. An efficient implementation can be found in Morard et al. (2013).

7.4 Attribute controlled reconstruction

As aforementioned, λ-flat zones are very useful in image partition, simplification and segmentation. However, it suffers from the well-known chaining effect of the single linkage clustering (Duda et al., 2000). That is, if two distinct image objects are separated by one or more transitions going step by step having a gray-level difference lower than λ, they will be merged in the same λ-flat zone.

To illustrate this effect, consider the toy example of Figure 7.5. The image contains two different objects (black square on the left and gray square on the right) connected by a segment with gradual gray-level transitions. Figure 7.5(b) shows the flat-zones of the image while Figure 7.5(c) shows the quasi-flat zones with a small λ value. Note that this propagation merges the two objects due to gradual gray-level transitions in the segment that connect them.

(a) original image (b) flat-zones (c) quasi-flat zones

Figure 7.5: The image contains two different objects (black square on the left and gray square on the right) connected by a segment with gradual gray-level transitions. (a) shows the flat-zones of the image while (b) shows the quasi-flat zones with a small λ value. Chaining effect due to small gray-level transitions in the segment that connects two different objects.

Several works try to restrict quasi-flat zones growth in order to prevent merging different regions. Hambrusch et al. (1994) propose a technique to limit the chaining effect by introducing an additional threshold that limits gray-level variation over the whole CC rather than just along connected paths. This relation is reflexive and symmetric, but not necessarily transitive, so it does not always lead to an image partition in the definition domain. Soille (2008) reviews several approaches and proposes a constrained connectivity called (λ, ω, β)-connectivity. In this approach, a succession of λ-flat zones is built with increasing slope parameter λ (up to a maximum λ_{max}), none of which may have gray-level difference greater than ω and connectivity index greater than β. This method has the advantage of providing a unique partition of the image domain, which is very difficult to achieve in any other way. This method was successfully applied to hierarchical image partition and simplification. Other solutions may include viscous propagations (Meyer and Vachier, 2002; Serra, 2005).

The main disadvantage of these approaches is the parameter tuning. With the aim of simplifying this selection, we propose an attribute controlled propagation based on increasing quasi-flat zones. It consists in evaluating attribute changes during region growing in order to select the appropriate partition. For a given attribute, no additional size parameter is required. In that sense, our method takes advantage of prior knowledge and intrinsic information of the image in order to define the best propagation.

The idea comes from the reconstruction of an object from a marker. Let us describe the problem with the

toy example of Figure 7.6. Consider a marker on the upper left corner of Figure 7.6(a) and its propagation by increasing λ-flat zones using 4-connected neighborhood. The propagation begins with $\lambda = 0$ (Figure 7.6(a)) and ends when propagation reaches the whole image at $\lambda = 5$ (Figure 7.6(f)).

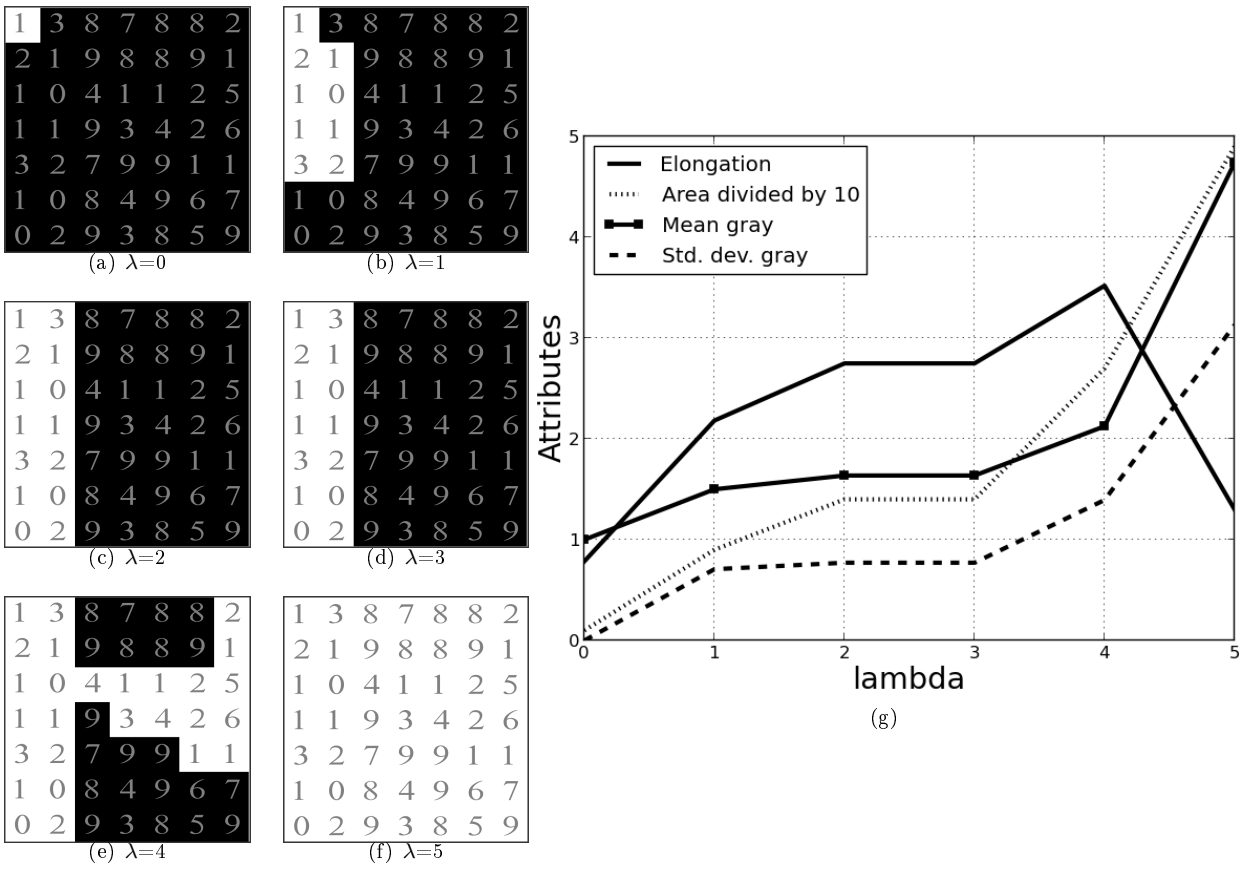


Figure 7.6: The idea of this propagation comes from the reconstruction of an object from a marker. Consider a marker on the upper left corner of (a) and its propagation by increasing λ-flat zones using 4-connected neighborhood. The propagation begins with $\lambda = 0$ (a) and ends when propagation reaches the whole image at $\lambda = 5$ (f). (g) presents the evolution of four attributes during this propagation: area $S(X)$, elongation $E(X)$, mean gray-level $\mu_I(X)$ and standard deviation of gray-level $\sigma_I(X)$.

From the image segmentation point of view, the question is: *when should propagation be stopped?*. Obviously, the answer is application dependent. Intuitively, the evolution of an attribute could be useful to make the decision. For example, Figure 7.6(g) presents the evolution of four attributes: area $S(X)$, elongation $E(X)$, mean gray-level $\mu_I(X)$ and standard deviation of gray-level $\sigma_I(X)$. We propose two criteria in order to stop the propagation:

- Attribute rupture: select the propagation such that the attribute change between two consecutive λ is maximum.
- Maximum attribute: select the propagation such that the attribute is maximum.

On the one hand, one can see between λ=3 and λ=4 that area increases up to 200% of its value (from 14 to 27 pixels). This great change is called an *attribute rupture*, and it can be a reason to stop the growing process. Another example occurs between λ=4 and λ=5, where ruptures are identified on $E(X)$, $\mu_I(X)$ and $\sigma_I(X)$. On the other hand, the maximum elongation occurs at λ=4. Note that for increasing attributes (*e.g.* area) the maximum attribute value always corresponds to the propagation on the whole image. Therefore, selecting the maximum attribute is only reasonable in the case of non-increasing attributes (*e.g.* elongation).

Based on Definition 7.3.1, let us introduce formal definitions for the set of increasing λ-flat zones:

Definition 7.4.1 *For all $x \in D$, let Λ_x be the set of increasing regions containing pixel x. For all $\lambda \in V$ and $j = [1, ..., n-1]$, we define $A_x(\lambda) \in \Lambda_x$ as the λ-flat zone of image I containing x:*

$$A_x(\lambda) = \{x\} \cup \{q | \exists \wp = (p_1 = x, ..., p_n = q) \text{ such that } |I(p_j) - I(p_{j+1})| \leq \lambda\} \tag{7.3}$$

In this section, λ-flat zones are arbitrarily used. However, this is not a restrictive choice since any other hierarchical partition can be used as well. Another application is presented later in Section 7.7, where a component tree is used.

Let us introduce formal definitions for attribute rupture and maximum attribute:

Definition 7.4.2 *Let $\Gamma(A_x(\lambda))$ be an attribute on the λ-flat zone of image I containing pixel x. For all $\lambda_i \in V$ and $i = [1, ..., n-1]$, we define λ_M and λ_R as the values for which the maximum attribute and the attribute rupture appear, respectively:*

$$\begin{aligned} \lambda_M &= argmax_{\lambda_i \in V} |\Gamma(A_x(\lambda_i))| \\ \lambda_R &= argmax_{\lambda_i \in V} |\Gamma(A_x(\lambda_i)) - \Gamma(A_x(\lambda_{i+1}))| \end{aligned} \tag{7.4}$$

In this controlled reconstruction, we only analyze one attribute at the same time. However, other statistics or combination of several attributes can be used as well, as it will be shown later in Section 7.7.2. Compared to other methods, the main advantages of our approach are: no size parameter is required in order to determine the adaptive region; it is a connected operator since the λ-flat zones do not create new contours on the image (Salembier and Serra, 1995; Salembier and Wilkinson, 2009); it is multi-scale since λ-flat zones size is not restricted; and it is auto-dual since bright, dark and intermediate gray-level regions are processed at the same time.

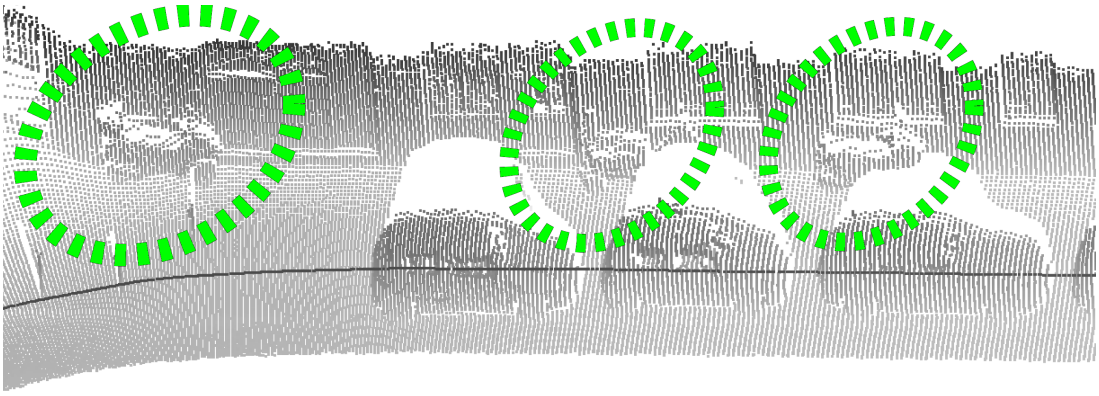
To illustrate the performance of this attribute controlled reconstruction, an application on image segmentation is presented: controlled propagation from markers in order to segment connected objects in elevation images. We present an automatic facade segmentation method developed as part of TerraMobilita project. For this purpose, a reconstruction controlled by elongation is applied.

It is noteworthy that facades are the highest and most elongated structures in the elevation image. Thus, a reconstruction controlled by elongation is applied. For this purpose, facade markers are defined based on height constraints. Let us concentrate on the reconstruction step since the marker selection is straightforward and it has been already explained in Chapter 5. We propose to apply a reconstruction from markers stopping when the elongation is maximum. Figure 7.7 illustrates the process on a urban scenario where three motorcycles are parked next to the facade. Figures 7.7(b) and 7.7(c) show pictures helpful to understand the scene. Figure 7.7(d) presents the elevation image and the markers located in the upper facade part. Figure 7.7(f) shows the elongation evolution using increasing λ values. Reconstruction at λ=13 is selected, which corresponds to the maximum elongation. Note that the maximum elongation (at λ=13) and the elongation rupture (at λ=14) are almost the same CC, thus this selection is not critical for this example. The reconstruction result on the elevation image is shown in Figure 7.7(e) while Figure 7.7(g) shows the result reprojected onto the 3D point cloud. One can see that the entire facade is correctly reconstructed without including connected motorcycles.

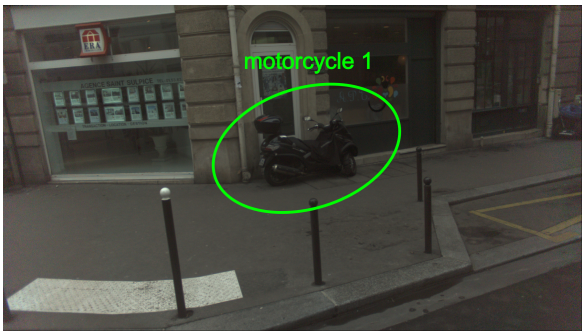
With respect to other approaches in which a parameter should be selected, our method only requires selecting an attribute, then the appropriate propagation is automatically selected. This is useful when segmenting objects with similar attributes on large databases. For example, facades are always the most elongated structures. Then, if different λ parameters are required to segment facades on different images or even different facades on the same image, our method will adapt the parameter to the best possible value.

7.5 Adaptive mathematical morphology

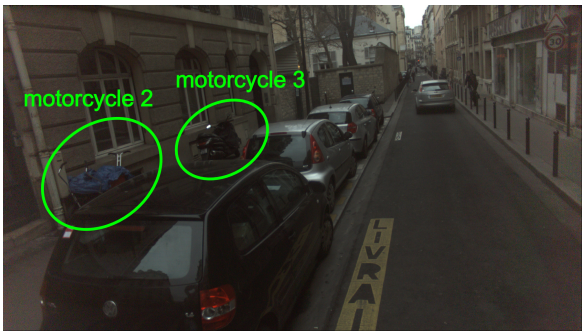
When using MM, square SE are preferred in most practical applications. However, several works remark the usefulness and necessity of adapting algorithms according to intrinsic variability and prior knowledge of the image (Maragos and Vachier, 2009). Adaptive SE are elegant processing techniques using non-fixed kernels. Such operators, firstly introduced by Gordon and Rangayyan (1984), vary their shape over the whole image taking into account local image features. Serra (1982) called them *structuring functions* and defined erosion and dilation with spatially-varying SE. In the literature, several works have been carried out with the aim of



(a) 3D point cloud showing three motorcycles parked next to the facade.



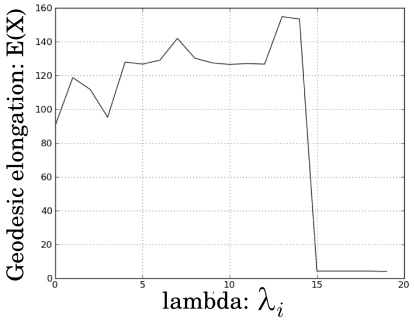
(b) Illustrative photo



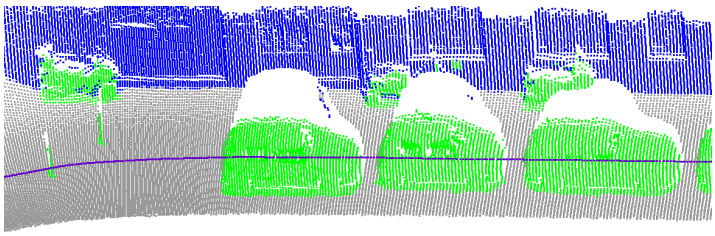
(c) Illustrative photo

(d) Elevation image: facade markers (black)

(e) Elevation image: facade reconstruction (black)



(f) Attribute evolution on the quasi-flat zones



(g) Segmented 3D point cloud. facades (blue), objects (green), ground (gray), vehicle trajectory (magenta).

Figure 7.7: Segmentation of connected objects using controlled propagation from markers. (b and c) pictures helpful to understand the scene. (d) elevation image and the facade markers. (f) shows the elongation evolution using increasing λ values. Reconstruction at λ=13 is selected, which corresponds to the maximum elongation. The reconstruction result is shown in (e) and (g).

exploiting image information in order to locally adapt SE shape and size. An overview on adaptive MM can be found in Maragos and Vachier (2009). Most works focus on filters that privilege smoothing in homogeneous regions while preserving edges as well as possible. With this idea, Perona and Malik (1990) proposed anisotropic filters that inhibit diffusion through strong gradients.

One of the first works using adaptive SE is due to Beucher (1987). He developed a traffic control application where the SE size depends on the perspective and varies linearly as the vertical position on the image on a video sequence. Later, Verly and Delanoy (1993) applied adaptive MM to range imagery to correct perspective distortions. Their approach consists in defining square SE such that their size depends on the distance between object and sensor. Shih and Cheng (2004) used simple and fast adaptive dilations with elliptic SE that varies its size and orientation according to local properties. Talbot and Appleton (2007) proposed a more sophisticated solution defining pixel connectivity by complete and incomplete paths. Pinoli and Debayle (2009) proposed a general adaptive neighborhood for MM as follows: given a criterion mapping h and a tolerance $m > 0$, at each point x an adaptive neighborhood is defined containing all points y such that $|h(y) - h(x)| < m$. Morard et al. (2011b) proposed adaptive SE based on a region growing process. These SE have a fixed size but they adapt their shape by choosing recursively homogeneous pixels with respect to the seed pixel. Angulo (2011) used the notion of counter-harmonic mean in order to propose bilateral filters which asymptotically correspond to spatially-variant morphological operators. More recently, Franchi and Angulo (2014) proposed a spatially-variant area opening in order to preserve contours according to a reference image. Its natural application domain is the video sequences. Among the different approaches in input-adaptive operators, morphological amoebas (Lerallut et al., 2007) appear as a promising solution. They adapt their shape according to a distance that depends on both the length and the gray-level differences on a neighborhood. This distance is used to define structuring elements $N(x) = \{y : d_\sigma(x, y) \leq r\}$ for each pixel on the input image. Because the amoeba distance is an increasing attribute, increasing r leads to an inclusion property useful to define operator pyramids (Serra and Salembier, 1993). Note that all those works are applied to MM, however they are useful to any other local operator such as convolution or non-linear filters.

Actually, if a given morphological processing consists in successive operators (*e.g.* an opening is an erosion followed by the reciprocal dilation), the SE should be the same for all of them in order to preserve mathematical properties of morphological filters, as proved by Roerdink (2009). Thus, adaptive SE are computed on a pilot image, the same for the whole process. This pilot image can be the original image or a filtered version of it in order to reduce noise impact in the SE shape.

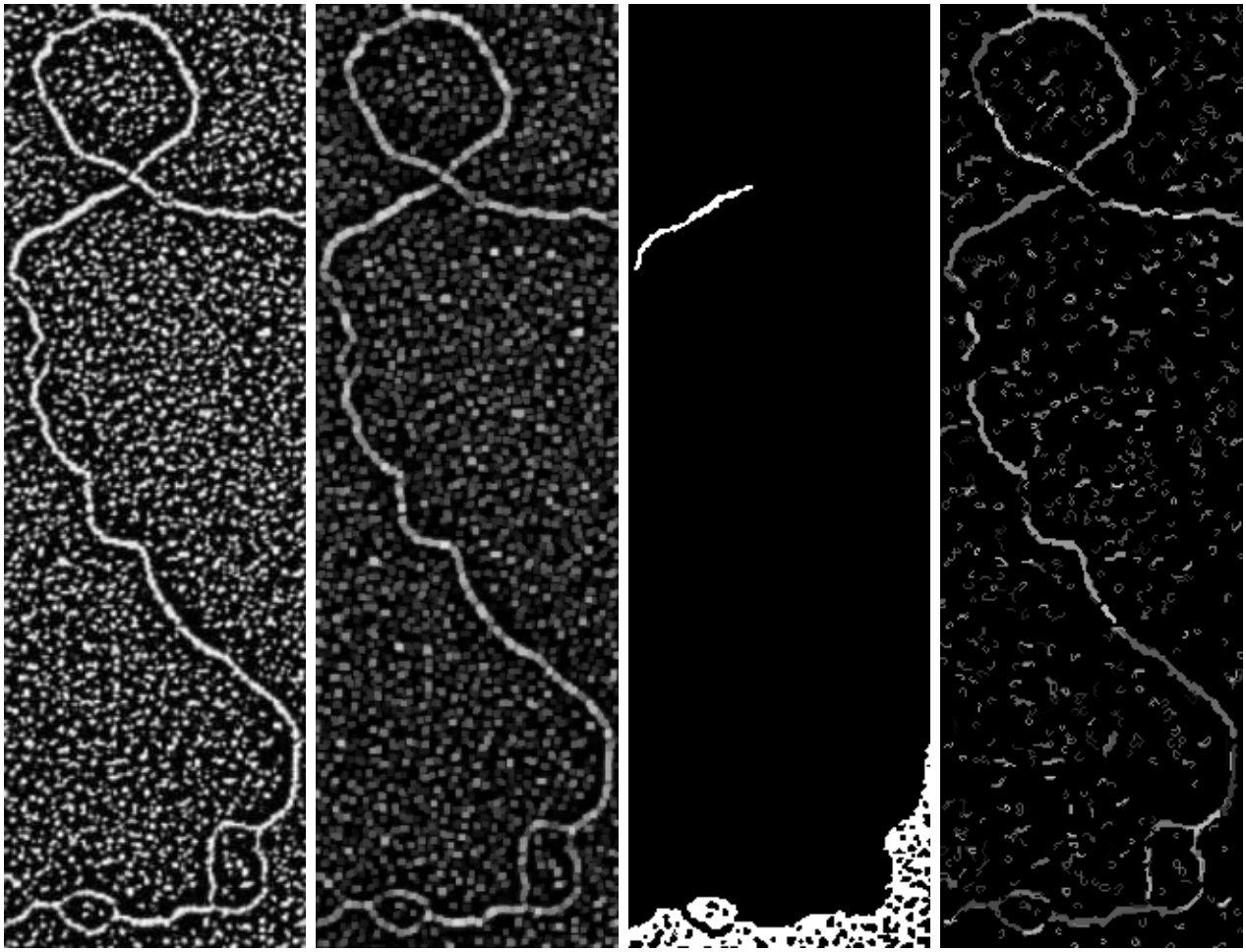
In this section, we propose to use our attribute controlled propagation (introduced in Section 7.4) on a pilot image in order to define input-adaptive SE for each pixel on the original image, similar to Lerallut et al. (2007); Grazzini and Soille (2008). Such adaptive SE are useful to filter structures according to a given attribute. For example, Figure 7.8 presents an opening with adaptive SE using the maximum elongation. Figure 7.8(c) illustrates the SE for two pixels in elongated (fiber) and non-elongated (background) regions. Figure 7.8(d) compares the result of an adaptive opening with respect to the classical one (Figure 7.8(b)). Note that elongated structures are preserved while non-elongated structures are merged with their neighborhood. Remaining small spurious regions may be filtered out using a simple area opening.

Figure 7.9(d) presents another example using the gray-level rupture to stop the propagation. This is useful to define SE containing pixels with similar gray-level. Figure 7.9(d) shows the SE for two different pixels in the image. Figure 7.9(e) presents the application of this adaptive SE as kernel for a non-linear filter, the median filter. Note that homogeneous regions are smoothed and high contrasted structures are preserved, as proven by the number of flat-zones of each filtered image (Figures 7.9(c) and 7.9(f)). Compared with amoebas and other similar works, our method does not require any additional size parameter since the SE only depends on the attribute selection and the input image.

7.6 Feature Extraction

In this section, we present another application to extract features from an image based on the shape of the input-adaptive SE. To the authors knowledge, this idea was firstly presented by Morard et al. (2011b), who propose an approach using region growing structuring elements (REGSE). For each pixel on the image, they define a neighborhood of N pixels minimizing a homogeneity function $\rho(x)$ (*e.g.* gray-level difference) between adjacent pixels. Then, they use the REGSE to compute shape features in the image. An advantage is that REGSE can follow any homogeneous structure, however it is not multi-scale because its size has to be exactly N pixels.

We propose a similar approach using our propagation method to define adaptive SE (Section 7.5). The



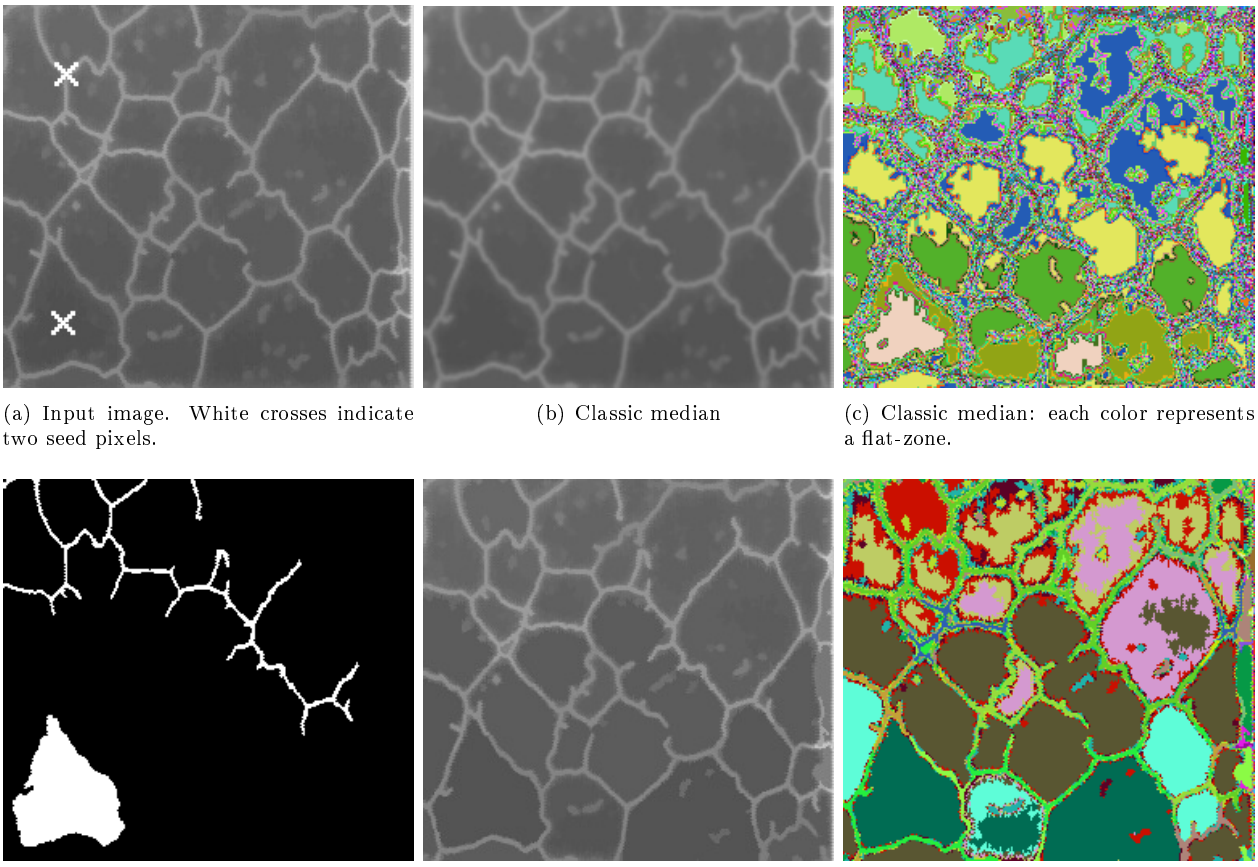
(a) Input image. (b) Classical opening of size 1. (c) Adaptive SE for non-elongated (background) and elongated (fiber) regions. (d) Adaptive opening.

Figure 7.8: Input-adaptive SE using the maximum elongation. In this case, the input and the pilot image are the same. (c) illustrates the SE for two pixels in elongated (fiber) and non-elongated (background) regions. (d) shows the result of an adaptive opening with respect to the classical one (b). Note that elongated structures are preserved while non-elongated structures are merged with their neighborhood. Remaining small spurious regions may be filtered out using a simple area opening.

main comparative advantage is that parameter N is not required, because it is adapted for each pixel during the propagation from it. In that sense, we use non-constant size SE that depends on the image intrinsic information. This is specially useful when the image contains objects at different scales. Additionally, remember that our propagation is a connected operator since λ-flat zones do not create new contours on the image. This is not true for REGSE, where region growing is forced to stop at N pixels.

Consider the two examples of Figure 7.10. From each pixel, we compute the adaptive SE using a propagation controlled by the maximal elongation. Each pixel on the output image contains the maximal elongation of its respective adaptive SE. It is noteworthy that the highest values in the feature image correspond to the most elongated structures. Moreover, this operator is auto-dual since brighter and darker structures are processed at the same time. See for example Figure 7.10(c), where several elongated vessels at different gray levels have been enhanced on the maximal elongation image (Figure 7.10(d)). If the user want to to favor a given gray level, the feature image can be weighted using the original input image.

Feature images are useful to assess features and segment structures of a given shape. Compared to geodesic thinnings (Morard et al., 2011a), that uses geodesic elongation as our method does, our approach has the following advantages: i) our feature image contains information about all objects in the scene, while geodesic



(a) Input image. White crosses indicate two seed pixels.

(b) Classic median

(c) Classic median: each color represents a flat-zone.

(d) Adaptive SE for the two seed pixels.

(e) Adaptive median

(f) Adaptive median: each color represents a flat-zone.

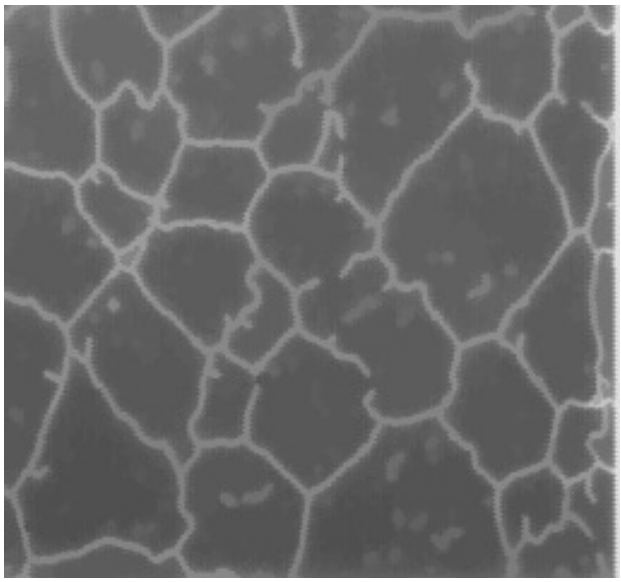
Figure 7.9: Input-adaptive SE using the gray-level rupture. In this case, the input and the pilot image are the same. This is useful to define SE containing pixels with similar gray-level. (d) shows the SE for two different pixels in the image. (e) presents the application of this adaptive SE as kernel for a non-linear filter, the median filter. Note that homogeneous regions are smoothed and high contrasted structures are preserved, as proven by the number of flat-zones of each filtered image (c) and (f). Compared with amoebas and other similar works, our method does not require any additional size parameter since the SE only depends on the attribute selection and the input image.

thinning must be computed every time in order to extract structures at different elongations; ii) our method, based on quasi-flat zones, deals with bright, dark and intermediate gray level regions at the same time whereas geodesic thinning focuses only on bright objects.

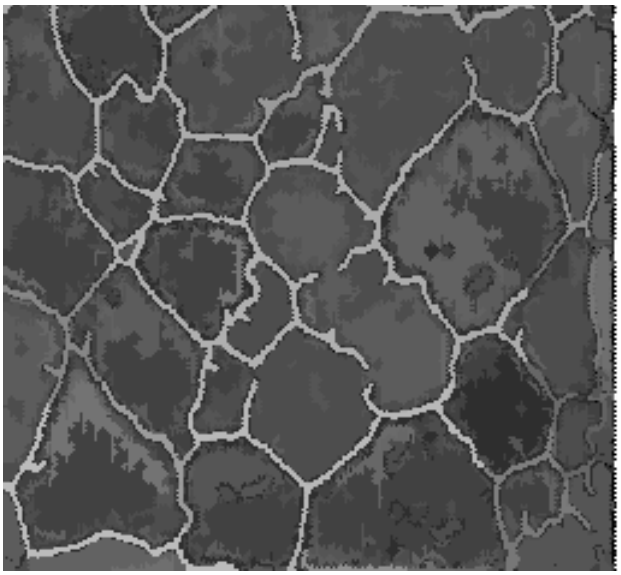
Consider for example Figure 7.11(a), where the aim is segmenting as much elongated structures as possible. Figures 7.11(c) and 7.11(d) present two geodesic thinnings at $E(x)$=11 and $E(x)$=20, respectively. Note that only bright objects have been extracted. Figures 7.11(e) and 7.11(f) present two simple thresholds on the maximal elongation image (Figure 7.11(b)) at these same values. It is noteworthy that our proposed operator is more appropriate to this segmentation task since black, white and gray elongated structures can be detected. Moreover, our feature image (Figure 7.11(b)) contains information about the elongation of all objects in the scene, proving the usefulness of our transformation for segmentation and classification tasks.

7.7 Attribute profiles and area-stable elongation

Filtering techniques, aiming at removing noise while preserving as much as possible the desired information, are often essential prior to segmentation. Several works aiming at filtering and segmenting objects based on attribute profiles can be found in the literature. Jones (1999) proposes connected filters using attributes signatures, *i.e.*

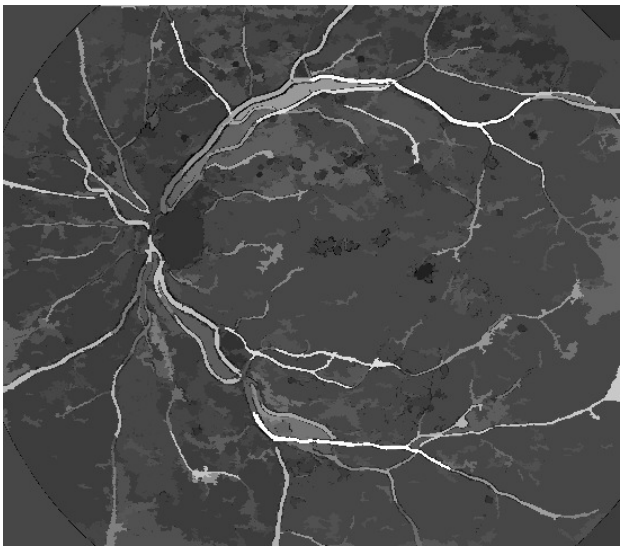


(a) Grains: original image.



(b) Grains: maximal elongation image by quasi-flat zones.

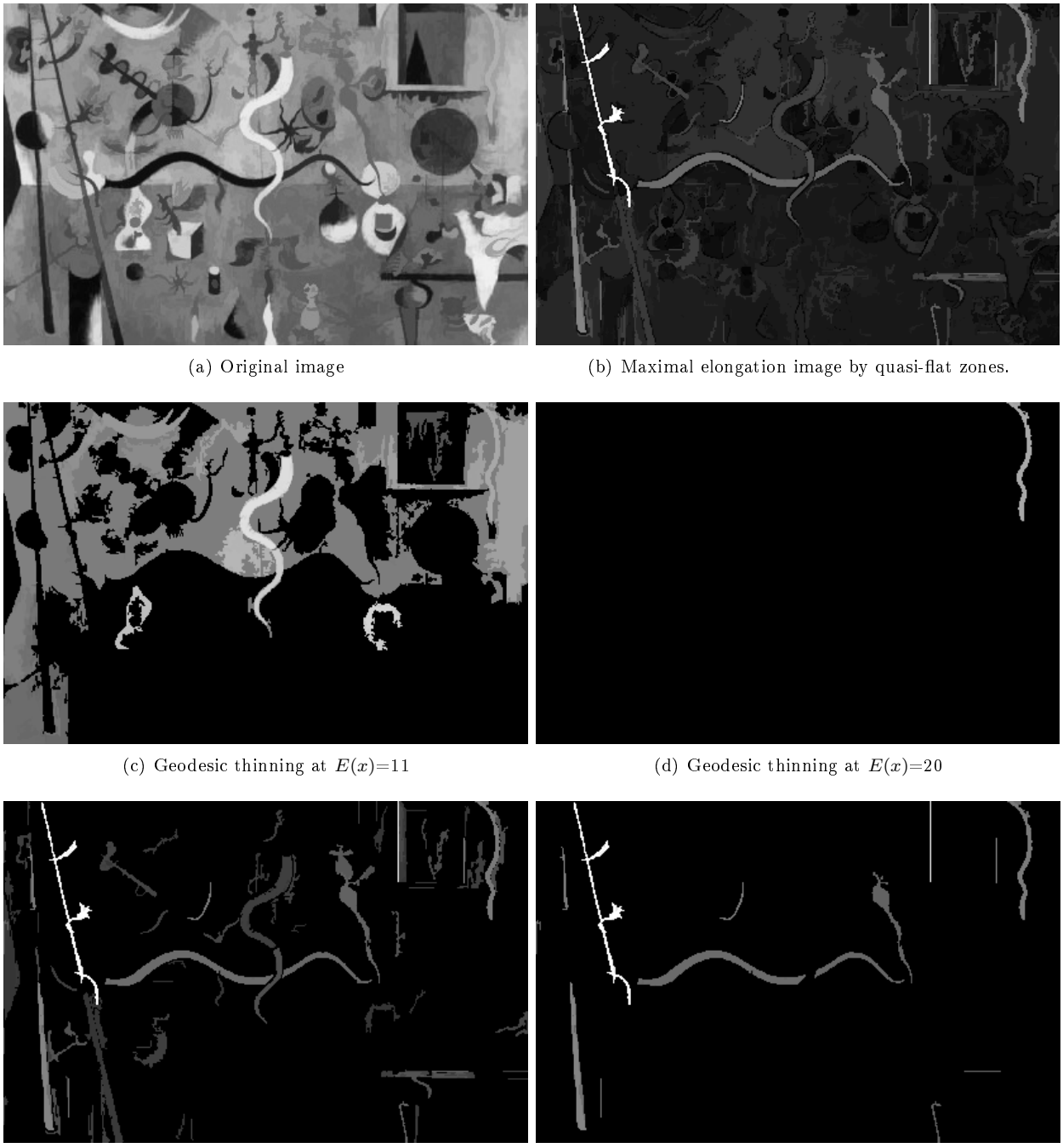
(c) Retina: original image.



(d) Retina: maximal elongation image by quasi-flat zones.

Figure 7.10: Feature images using input-adaptive SE. Quasi-flat zones propagation controlled by the maximal elongation. Each pixel on the output image contains the maximal elongation of its respective adaptive SE. It is noteworthy that the highest values in the feature image correspond to the most elongated structures. Moreover, this operator is auto-dual since brighter and darker structures are processed at the same time. See for example (c), where several elongated vessels at different gray levels have been enhanced on the maximal elongation image (d). If the user want to to favor a given gray level, the feature image can be weighted using the original input image.

the evolution of an attribute on the component tree. He has successfully applied his method to the segmentation of wood micro-graphs. Pesaresi and Benediktsson (2001) introduce morphological profiles using the derivative of the residues from openings and closings by reconstruction. Their method is well suited for images with low contrast and low resolution. However, the maximal residue may not be the best segmentation choice. Moreover, the computational cost increases when processing large and homogeneous images. Beucher (2007) proposes the analysis of the residue through successive morphological operations. This evolution over each image pixel leads to interesting transformations such as ultimate openings and quasi-distance functions. Ouzounis et al. (2012) propose differential area profiles for efficient point-based multi-scale feature extraction in pattern analysis and



(a) Original image

(b) Maximal elongation image by quasi-flat zones.

(c) Geodesic thinning at $E(x)$=11

(d) Geodesic thinning at $E(x)$=20

(e) Maximal elongation image. Threshold at $E(x)$=11

(f) Maximal elongation image. Threshold at $E(x)$=20

Figure 7.11: Segmentation of elongated structures at different thresholds using geodesic thinnings (Morard et al., 2011a) and thresholding on the maximal elongation image computed using quasi-flat zones. It is noteworthy that our proposed operator is able to segment black, white and gray elongated structures. (c) and (d) present two geodesic thinnings at $E(x)$=11 and $E(x)$=20, respectively. Note that only bright objects have been extracted. (e) and (f) present two simple thresholds on the maximal elongation image (b) at these same values. It is noteworthy that our proposed operator is more appropriate to this segmentation task since black, white and gray elongated structures can be detected. Moreover, our feature image (b) contains information about the elongation of all objects in the scene, proving the usefulness of our transformation for segmentation and classification tasks.

image segmentation.

In this section, we propose a method to segment elongated objects based on the analysis of the attribute profile over the threshold decomposition of an image. We define a new attribute, called area-stable elongation, that combines elongation and area stability. In our experiments, we analyze important events in the evolution of this attribute and we show its efficiency in segmenting elongated objects while filtering out noisy structures. An application aiming at segmenting elongated cells (melanocytes) in multiphoton fluorescence microscopy images of engineered skin is presented. Another application on the segmentation of facades from 3D urban data is detailed in Chapter 5.

7.7.1 Maximally Stable Extremal Regions (MSER)

The Maximally Stable Extremal Regions (MSER) method, proposed by Matas et al. (2004), is a well-known region detector. MSER are invariant to affine transformations of both intensity and image coordinates. They have a high repeatability and can be run in linear time with respect to the number of pixels in the image (Nistér and Stewénius, 2008). However, the parameter selection remains its major drawback. Even when default parameters perform well in many applications, some heuristics need to be applied in order to yield appropriate regions. Moreover, MSER favors round regions, as proved by Kimmel et al. (2011), making it unsuitable to detect irregular shapes such as elongated objects.

Forssen and Lowe (2007) compute SIFT descriptors on each MSER region in order to extract image features. This approach is proven to be robust to illumination changes and nearby occlusions. They also proposed a pyramidal decomposition of the image in order to get scale invariability. The authors also suggested the use of MSER for image segmentation. Forssen (2007) extends the MSER concept to color images and Litman et al. (2012) define stable volumetric features in deformable shapes.

Using threshold decomposition, the attribute profile (Section 7.3.2) can be used to characterize and to filter structures on the image. A simple but interesting attribute is the area $S(X_t)$. When used to suppress small CC, it leads to the definition of area opening (Vincent, 1994). Since $S(X_t)$ is increasing, events in the area profile are analyzed instead of its global maximum. For example, great changes in area are probably related to the union of different objects while small ones are related to area stable regions, as those detected by the MSER method (Matas et al., 2004).

The area stability $\Psi(X_t)$ of the region X_t is defined as the ratio between its area $S(X_t)$ and its area variation $dS(X_t)/dt$, as shown in Equation (7.5):

$$\Psi(X_t) = \frac{S(X_t)}{dS(X_t)/dt} \tag{7.5}$$

A MSER is a connected component X_t with maximal area stability. In the original proposition, every local maximum is detected. Thus it is possible to have nested regions and some heuristics are required to select only the most important peaks.

7.7.2 A new attribute: Area-stable elongation

Favoring regular (round) regions is one of the main limitations of MSER, as proved by Kimmel et al. (2011). Thus, it is not suitable to detect irregular shapes, as elongated objects. In order to detect elongated objects taking into account their area stability, we propose a new attribute $\Phi(X_t)$, called *area-stable elongation*. This attribute combines area stability $\Psi(X_t)$ and elongation $E(X_t)$, as defined in Equation (7.6):

$$\Phi(X_t) = \Psi(X_t)E(X_t) = \frac{S(X_t)}{dS(X_t)/dt}\frac{\pi}{4}\frac{L^2(X_t)}{S(X_t)} = \frac{\pi}{4}\frac{L^2(X_t)}{dS(X_t)/dt} \tag{7.6}$$

The area-stable elongation $\Phi(X_t)$ is affine-invariant since $\Psi(X_t)$ and $E(X_t)$ are preserved under affine transformation of intensity and image coordinates, as stated by Matas et al. (2004) and Forssen and Lowe (2007). However, area-stable elongated regions are not invariant to blur. If blurring invariance is required, *e.g.* for a matching application, two solutions are possible, as proposed by Kimmel et al. (2011): i) weighting the stability function by the gradient magnitude along its boundary; ii) preprocessing the image with a deblurring filter.

The maxima of $\Phi(X_t)$ represent area stable regions with significant elongation. We propose to build a feature image using the maximal area-stable elongation $\Phi(X_t)$, which implies: i) the feature image is a partition of the

space, useful for segmentation; and, ii) each pixel contains information about shape and area stability of its neighborhood, which can be exploitable using prior knowledge.

Let us explain this new attribute with a toy example. Consider the 9×9 image of Figure 7.12. For this example, we have approximated the euclidean distance on the 8-connectivity grid, *i.e.* the geodesic diameter of a pixel is equal to 1, the distance between horizontal and vertical neighbors is equal to 1, and the distance between diagonal neighbors is equal to $\sqrt{2}$. This toy image contains 4 gray-levels enumerated from t_0 to t_3, and 6 CC enumerated from A to F. Figure 7.12(e) presents the component tree, where $S(X)$, $E(X)$, $\Psi(X)$ and $\Phi(X)$ are the area, the elongation, the area stability and the area-stable elongation of a given component X, respectively. Note that the stability for the background (object A) is not defined since it is the root of the component tree. The component tree contains two branches. Supposing we aim at segmenting object C from the left branch, and object E+F from the right one. Let us analyze each case separately.

First, object C is an elongated object nested on a spurious elongated structure B. Analyzing the elongation profile, we can see that object B ($E(X_B)$=6.10) is more elongated than object C ($E(X_C)$=4.60), as shown in the maximal elongation image of Figure 7.12(b). However, the stability of region C ($\Psi(X_C)$=2.43) is higher than that of region B ($\Psi(X_B)$=1.27), as shown in Figure 7.12(c). Combining these two attributes, region C ($\Phi(X_C)$=11.18) has a higher area-stable elongation than region B ($\Phi(X_B)$=7.71), as shown in Figure 7.12(d).

Second, object E is an elongated object that includes another elongated object F. Analyzing the elongation profile, we can see that object E+F ($E(X_E)$=8.43) is more elongated than the single object F ($E(X_F)$=3.14), as shown in the maximal elongation image of Figure 7.12(b). Since their area stabilities are similar ($\Psi(X_E)$=1.50 and $\Psi(X_F)$=1.80, as shown in Figure 7.12(c)), the highest area-stable elongation is obtained for the union of these two objects ($\Phi(X_E)$=12.65), as shown in Figure 7.12(d).

It is noteworthy that applying a simple threshold (*e.g.* $\Phi(X_t) \geq 8$) in the maximal area-stable elongation image (Figure 7.12(d)), objects C and E+F are correctly segmented, which is not possible on the original image (Figure 7.12(a)) nor on the other two feature images (Figures 7.12(b) and 7.12(c)).

Figure 7.13 illustrates the behavior of our method on a real DNA image. The goal is to segment the elongated and bright fiber from the noisy background. Figure 7.13(b) shows the maximal elongation image, where objects of an elongated shape are highlighted. However, spurious objects can be merged at low levels resulting in CC with high feature value, such as the porous structure in the center of the image. The maximal area stability (Figure 7.13(c)) keeps also many noisy and non elongated structures in the background. Finally, Figure 7.13(d) shows the area-stable elongation image, where most of noisy structures have been eliminated due to their low stability.

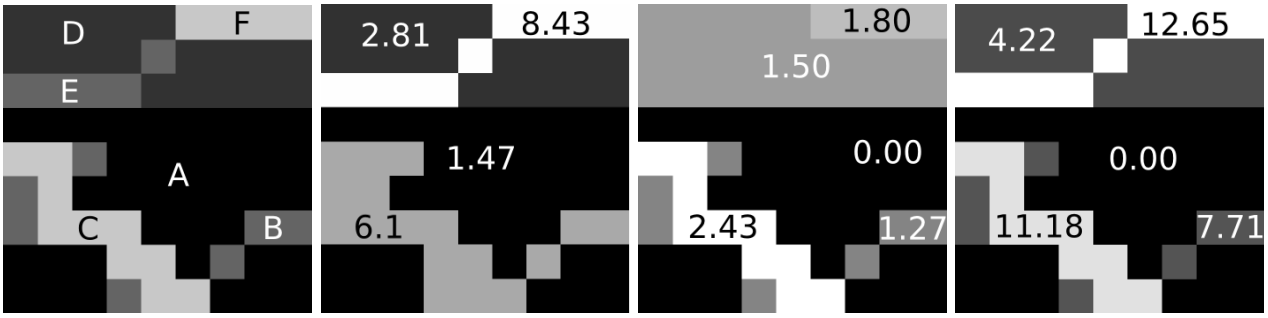
7.7.3 Application: segmentation of elongated cells

To illustrate the performance of our method, we apply it to segment elongated cells in multiphoton fluorescence microscopy images. Images correspond to reconstructed skin used in cosmetic research in applications such as screening of de-pigmenting and pro-pigmenting agents (Figure 7.14(a)). This model contains two types of cells: keratinocytes and melanocytes. The latter are dendritic cells, more elongated and brighter than keratinocytes. An accurate segmentation of melanocytes becomes crucial in order to quantify the melanin in the skin. This value is used to assess the efficiency of the cosmetic ingredient. Our goal here is to segment melanocytes, which appear as bright elongated structures.

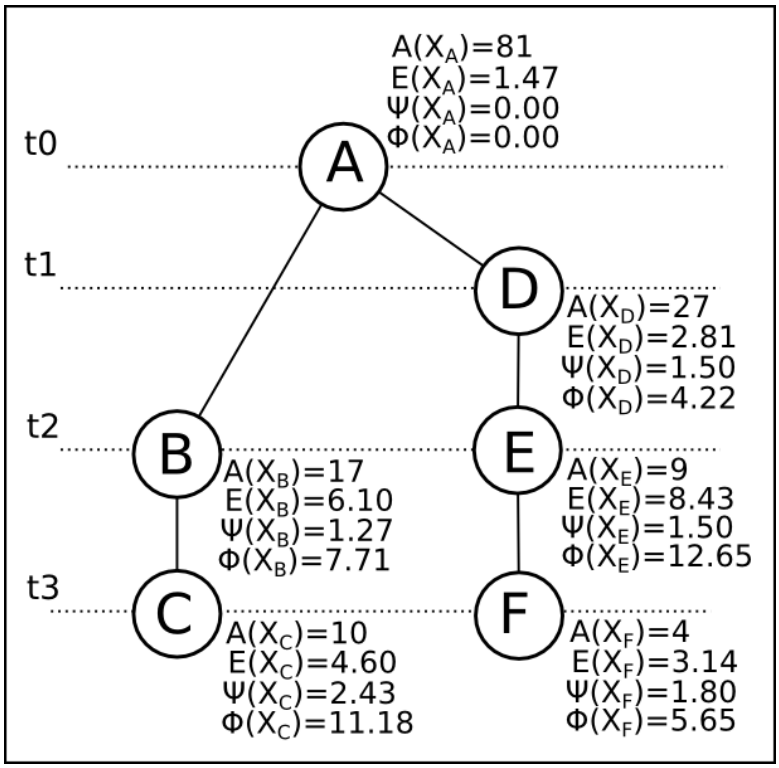
Segmenting these images with standard methods may fail since melanocytes are low contrasted and noisy, as shown in Figure 7.14(a). A first simple solution may consist in applying automatic thresholding, *e.g.* Otsu method (Otsu, 1979). However, it does not work because foreground and background gray-distributions overlap, as shown in the histogram of Figure 7.14(c). Thus, cells and background are not separable with a global threshold.

In this application, we propose a segmentation method using the component tree in order to solve the problem of low contrasted cells. Besides, the use of the area-stable elongation introduces shape prior knowledge and offers robustness to noise. In such a case, each cell can be segmented if it appears in the component tree, even if its gray-level is much lower than that for other cells in the image. Moreover, thanks to prior knowledge about melanocyte shape, the result is improved, justifying the use of our proposed methodology.

In our experiments we have 8 manually annotated images of 511 × 511 pixels each. The spatial resolution is equal to 0.26 μm/pixel. The ground truth definition has been carried out by experts from L'Oréal Research and Innovation (Serna et al., 2014a). Classical Precision (P), Recall (R) and $f_{\text{mean}} = (2 \times P \times R)/(P + R)$ statistics are computed in order to evaluate our results. The recall (or completeness) is defined as the number of correctly segmented pixels divided by the number of pixels marked in the ground truth. The precision (or



(a) Toy image (b) Max. elongation (c) Max. area stability (d) Max. area-stable elongation

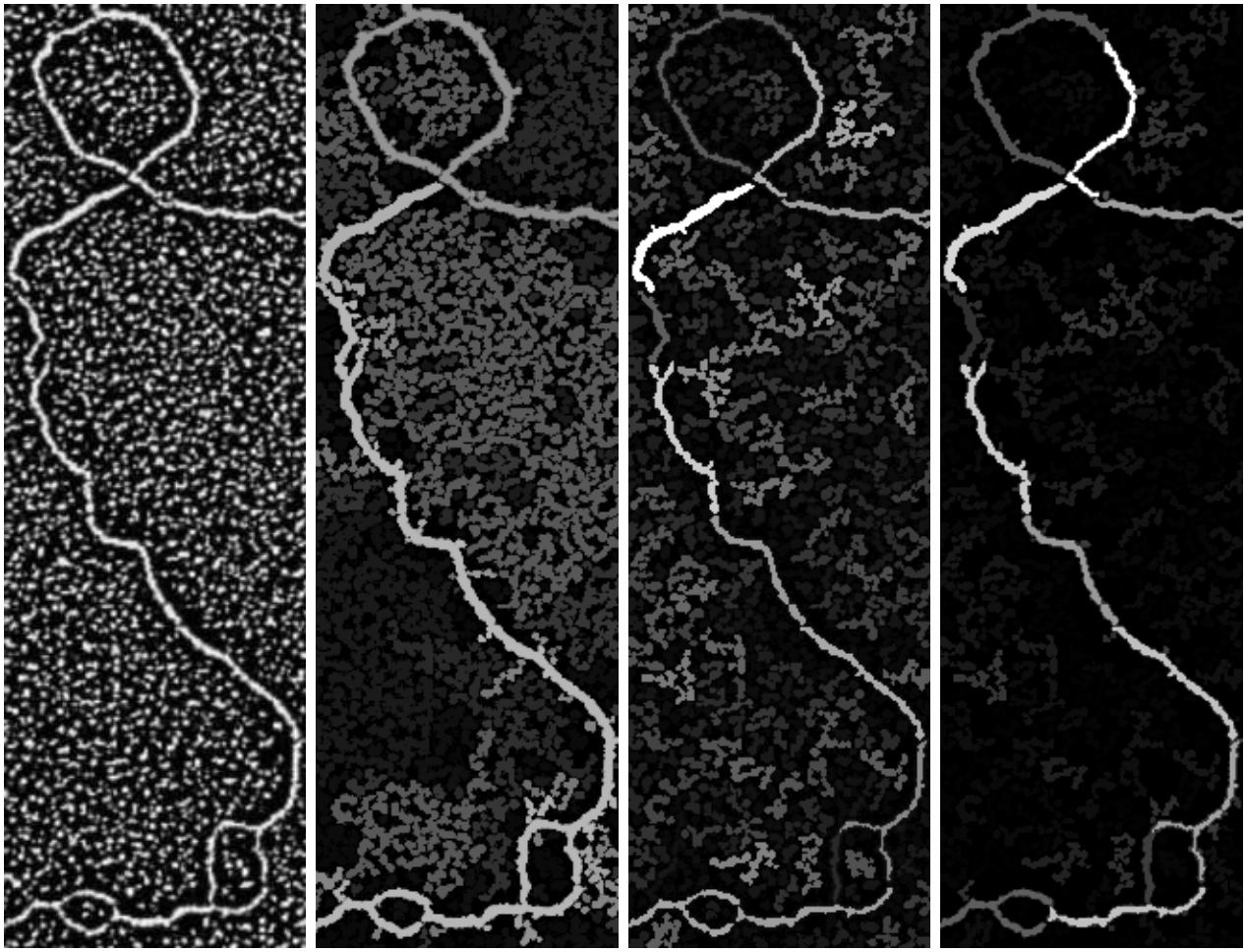


(e) Component tree

Figure 7.12: Toy example: maximal attributes images and component tree. For this example, we have approximated the euclidean distance on the 8-connectivity grid. This toy image contains 4 gray-levels enumerated from t_0 to t_3, and 6 CC enumerated from A to F. (e) presents the component tree, where $S(X)$, $E(X)$, $\Psi(X)$ and $\Phi(X)$ are the area, the elongation, the area stability and the area-stable elongation of a given component X, respectively. Note that the stability for the background (object A) is not defined since it is the root of the component tree. The component tree contains two branches. Supposing we aim at segmenting object C from the left branch, and object E+F from the right one.

correctness) is defined as the number of correctly segmented pixels divided by the total number of segmented pixels.

To exemplify our method, let us analyze the attribute profile for a single pixel belonging to a melanocyte, called seed pixel and marked with a red x in Figure 7.15(a). Figure 7.15(b) presents the ground truth provided by an expert. Figure 7.15(c) shows four attribute profiles: area $S(X_t)$, elongation $E(X_t)$, area stability $\Psi(X_t)$ and area-stable elongation $\Phi(X_t)$. For visualization purposes, each attribute has been normalized dividing by its maximum value to be in the range $[0, 1]$. Additionally, the f_{mean} is plotted in order to define the best possible



(a) DNA image

(b) Maximal elongation image: $\max\{E(X_t)\}$

(c) Maximal area stability image: $\max\{\Psi(X_t)\}$

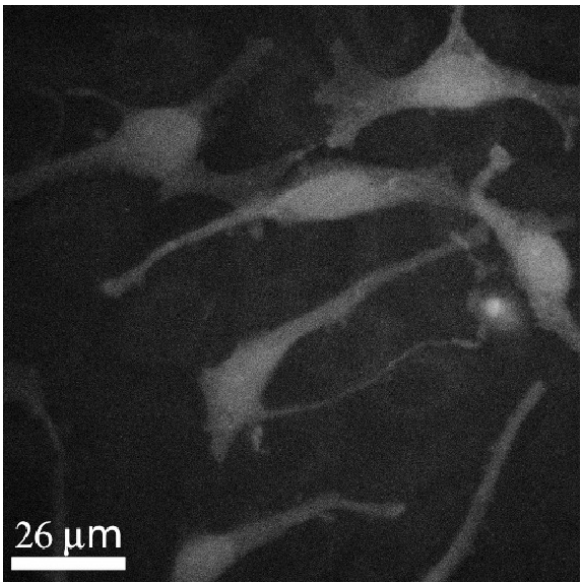
(d) Maximal area-stable elongation image: $\max\{\Phi(X_t)\}$

Figure 7.13: Maximal elongation, area stability and area-stable elongation on a real DNA image. The goal is to segment the elongated and bright fiber from the noisy background. (b) shows the maximal elongation image, where objects of an elongated shape are highlighted. However, spurious objects can be merged at low levels resulting in CC with high feature value, such as the porous structure in the center of the image. The maximal area stability (c) keeps also many noisy and non elongated structures in the background. Finally, (d) shows the area-stable elongation image, where most of noisy structures have been eliminated due to their low stability.

segmentation for this cell. Figures 7.15(d) to 7.15(i) show the evolution of the CC X_t containing the seed pixel.

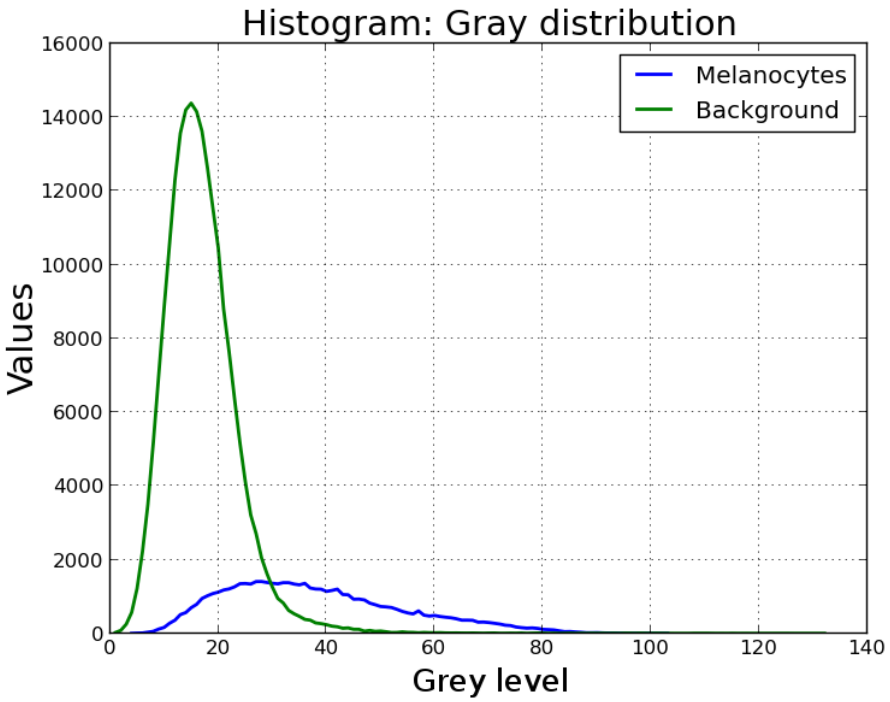
Segmentation methods using threshold decomposition are based on the hypothesis that objects of interest exist at some level of the tree. In our example, the best possible segmentation corresponds to $X_{t=34}$, for which the highest f_{mean} is obtained. Other good segmentations are in the range $X_{t\in[34,30]}$. The whole melanocyte is not retrieved for $X_{t>34}$ and it is merged with other structures for $X_{t<30}$.

Let us analyze each attribute profile, starting with $S(X_t)$. Based on prior knowledge about melanocytes size, attributes for $t < 15$ are not analyzed since they correspond to structures bigger than 75% of the whole image. Analyzing $\Psi(X_t)$, its global maximum represents the most stable region $X_{t=42}$. This is an area-stable and round region but useless in such a case since it does not match the entire melanocyte. Another interesting attribute is $E(X_t)$ since melanocytes are long and thin. Its global maximum corresponds to a CC merging three different objects $X_{t=28}$. From the area-stability point of view, this region is not stable because it is generated merging three different objects in a small range $t \in [30, 28]$. Finally, the global maximum of the area-stable elongation $\Phi(X_t)$ appears at $X_{t=34}$, which is the best segmentation according to f_{mean}.



(a) 511 × 511 pixels (resolution 0.26 μm/pixel).

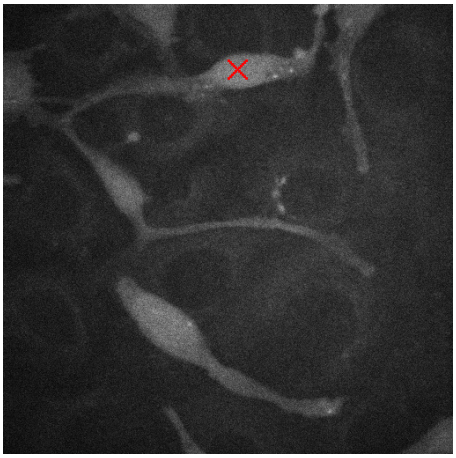
(b) Ground truth (melanocytes)



(c) Histogram of melanocytes and background.

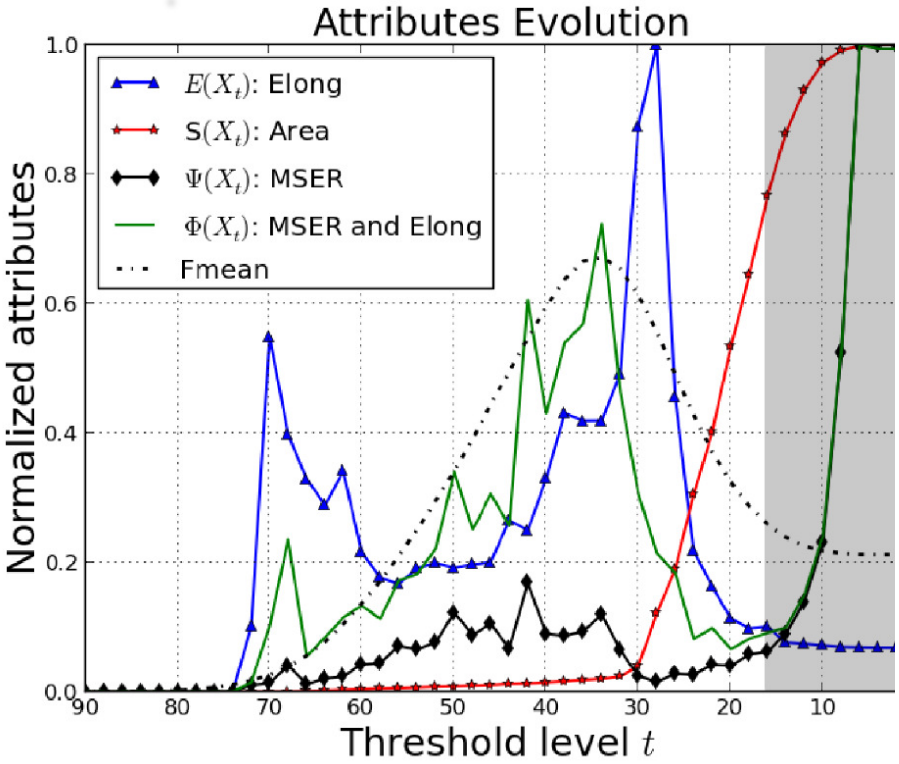
Figure 7.14: Foreground and background gray distributions on a multiphoton image of engineered skin containing keratinocytes and melanocytes. Segmenting these images with standard methods may fail since melanocytes are low contrasted and noisy, as shown in (a). A first simple solution may consist in applying automatic thresholding, *e.g.* Otsu method (Otsu, 1979). However, it does not work because foreground and background gray-distributions overlap, as shown in the histogram of (c). Thus, cells and background are not separable with a global threshold.

Figures 7.16 and 7.17 present two experimental results. Figures 7.16(a) and 7.17(a) show the two input images with their corresponding manual annotations in Figures 7.16(b) and 7.17(b). Figures 7.16(c) and 7.17(c) present the $\max\{E(X_t)\}$ images. Note that all melanocytes present a significant elongation, however some post

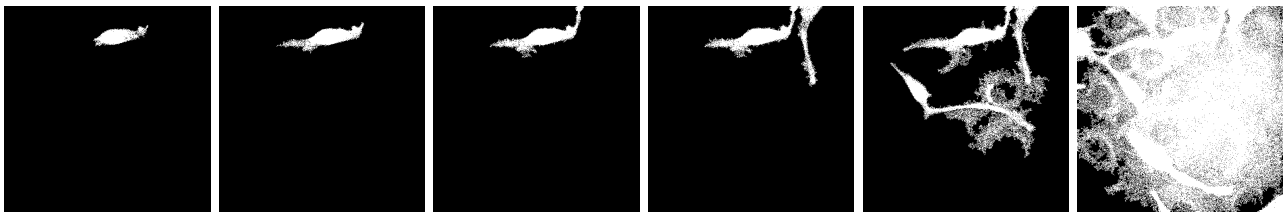


(a) Multiphoton image and a seed pixel.

(b) Ground truth.



(c) Attribute profiles. $X_{t<15}$ is not considered because $S(X_t) > 0.8$



(d) $X_{t=42}$ (e) $X_{t=34}$ (f) $X_{t=31}$ (g) $X_{t=30}$ (h) $X_{t=28}$ (i) $X_{t=16}$

Figure 7.15: Attribute profiles for pixel marked with a red x in the input image. From (d) to (i): CC of X_t containing pixel x, for different t values.

processing is required in order to eliminate porous structures on the background. Most noisy regions are not area-stable, then the area-stable elongation $\Phi(X_t)$ appears suitable for the segmentation of this kind of objects, as shown in Figures 7.16(d) and 7.17(d). This example demonstrates the use of our area-stable elongation in order to enhance elongated objects with respect to a noisy background. Using this feature image, the melanocyte segmentation becomes an easy task.

A simple three-fold segmentation algorithm is used for this purpose: i) characterization: a feature image is computed using the maximal area-stable elongation $\max\{\Phi(X_t)\}$; ii) filtering small objects: in the feature image, small regions (smaller than 500 pixels) are eliminated using an area opening followed by an area closing. This parameter is not critical since the smallest cell in the database is approximately 3000 pixels size; finally, iii) filtering objects with low attribute value: a simple threshold removes structures with low area-stable elongation. In our experiments we have used a threshold equal to 11 for all images. However, this parameter is not critical since several values produce similar results, as shown in the overall sensibility curve of Figure 7.18. It is noteworthy that thresholds between 7 and 16 produce an overall f_{mean} over 70%.

Table 7.1 presents quantitative results and a comparison with respect to the classical MSER (Matas et al., 2004). MSER regions have been computed using the algorithm directly provided by the authors (Mikolajczyk et al., 2005). The MSER parameters have been consistently set with those used by our method, *i.e.* the minimal MSER area has been set to 500 pixels and the threshold decomposition has been carried out for all gray-levels (one by one) stopping when the object area is greater than 75% of the whole image. Other MSER parameters such as relative area and relative margins have been kept to their default values 0.010 and `false`, respectively. Setting up these two parameters is not intuitive and global improvements have not been obtained in our tests. Using our method, this kind of parameters is not required, which is a clear advantage. MSER results may be also improved using some preprocessing step. However, it would require the setting up and the selection of the appropriate filter to do it. Another advantage of our method is that preprocessing is not used since the noise robustness is included in the area-stable elongation itself, as aforementioned in Section 7.7.2.

Table 7.1: Melanocyte segmentation: comparison with respect to MSER (Mikolajczyk et al., 2005). In each column, numbers on the left correspond to the proposed method, and numbers between parentheses to MSER.

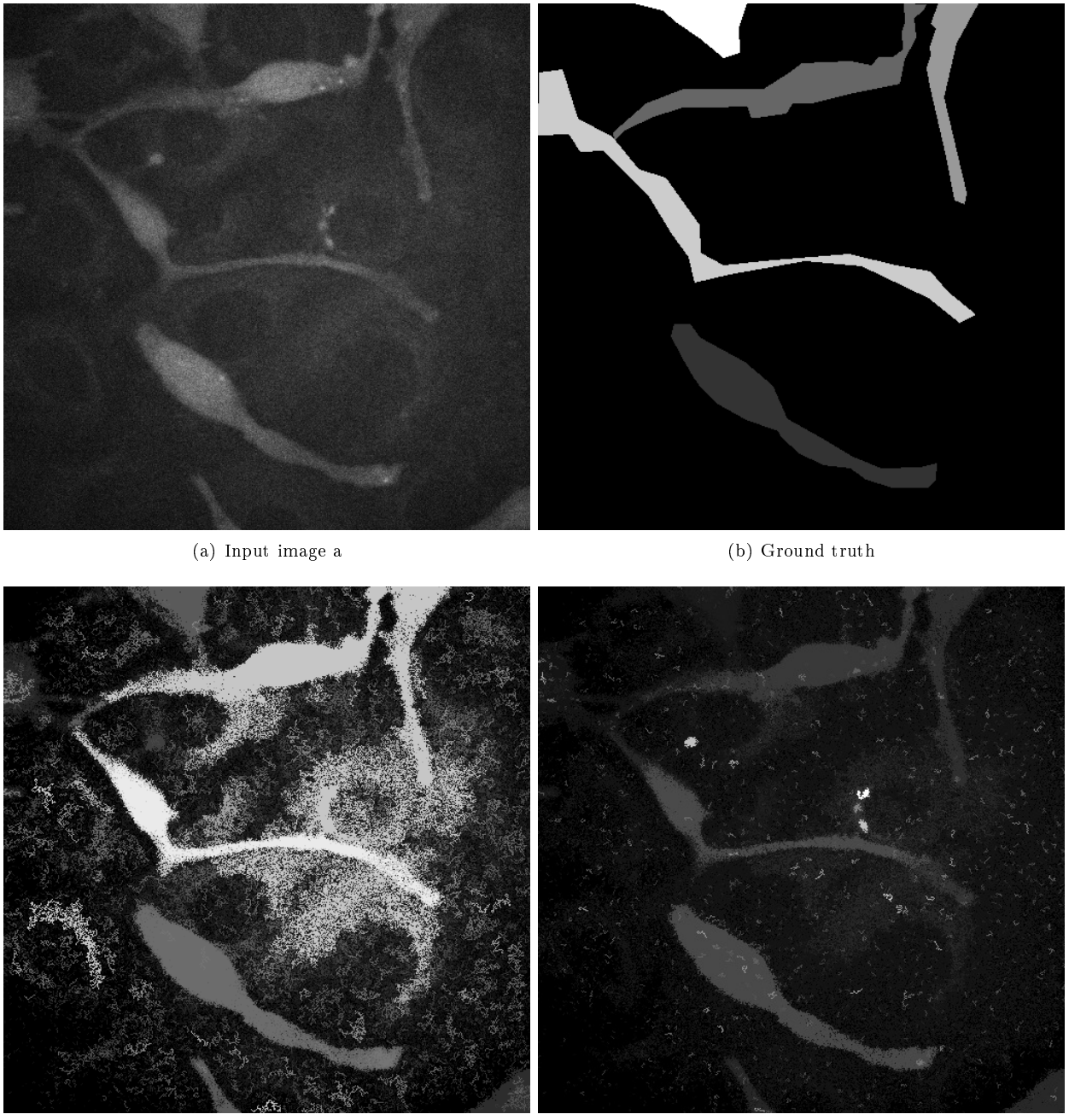
Image	Precision %	Recall %	f_{mean} %
a	84.0 (83.0)	61.0 (36.0)	71.0 (51.0)
b	74.0 (89.0)	84.0 (42.0)	78.0 (57.0)
c	71.0 (52.0)	84.0 (38.0)	77.0 (44.0)
d	78.0 (84.0)	78.0 (53.0)	78.0 (65.0)
e	85.0 (73.0)	90.0 (40.0)	87.0 (52.0)
f	84.0 (72.0)	68.0 (48.0)	75.0 (57.0)
g	62.0 (52.0)	92.0 (26.0)	74.0 (34.0)
h	83.0 (50.0)	86.0 (41.0)	84.0 (45.0)
Overall:	**78.0** (69.0)	**80.0** (41.0)	**78.0** (51.0)

Figures 7.19 and 7.20 present our experimental results showing the input image, the ground truth, the MSER result and our segmentation result. It is noteworthy melanocytes are correctly segmented by our method in most cases. Some problems are shown in Figure 7.19(c) where a clearly non-elongated melanocyte in the upper left part of the image has not been segmented, and in Figure 7.20(f) where a low contrasted melanocyte has been wrongly merged with the background. Note that our method presents much better results than MSER for all images. As aforementioned, MSER favors round and regular regions. Thus, only a partial segmentation is possible. Actually, MSER corresponds in several cases to the cell nuclei.

7.8 Conclusions

In this chapter, several methodological contributions to mathematical morphology have been presented. We have developed powerful attribute-based operators useful in a wide range of applications such as attribute controlled reconstruction, adaptive mathematical morphology, feature extraction, filtering and segmentation.

First, we have presented a reconstruction controlled by the evolution of a given attribute. The idea comes from the propagation from markers over increasing quasi-flat zones. We have shown that this method is a



(a) Input image a

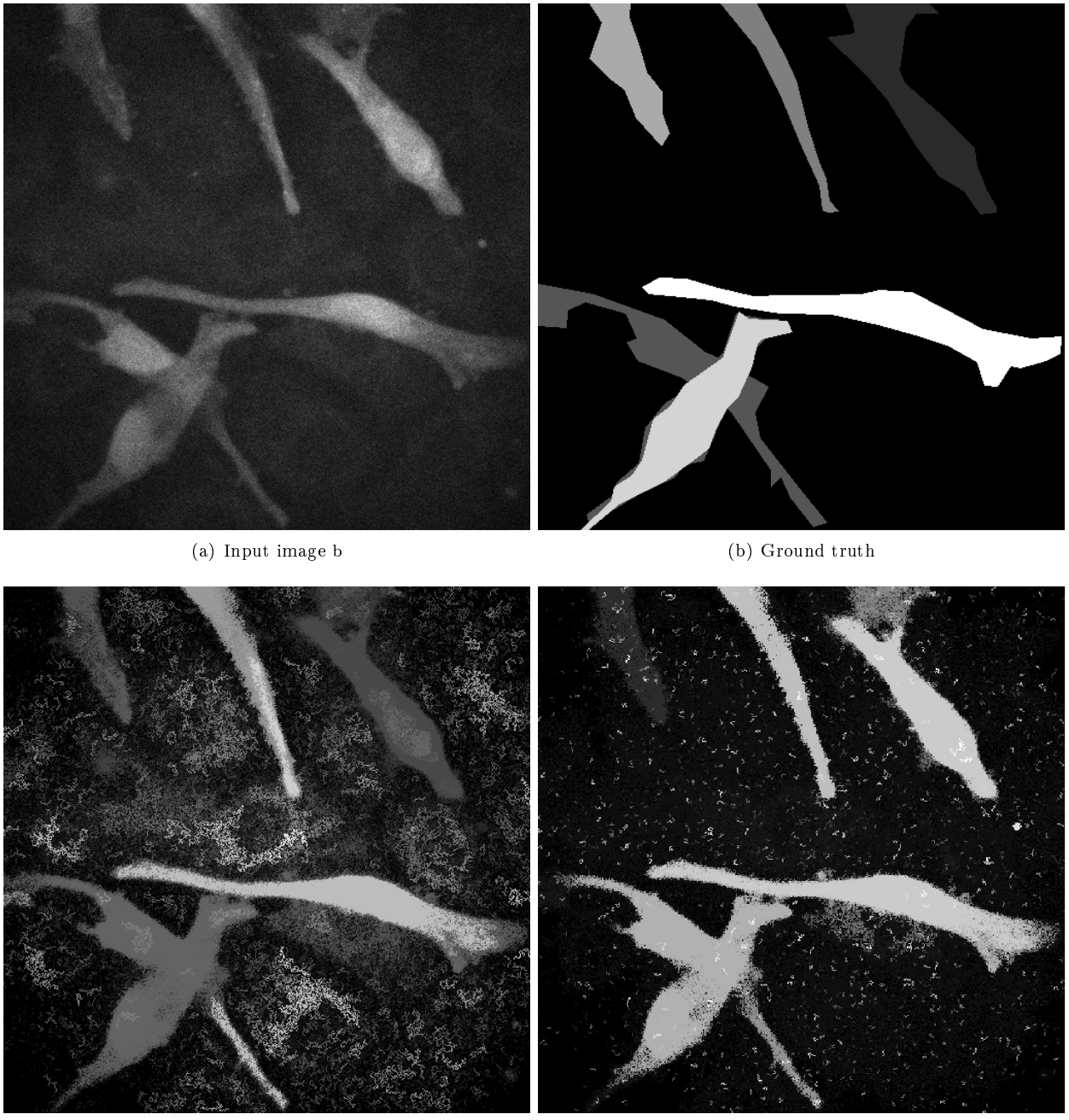
(b) Ground truth

(c) Maximal elongation image: $\max\{E(X_t)\}$

(d) Maximal area-sable elongation image: $\max\{\Phi(X_t)\}$

Figure 7.16: Feature images using the global maximum in the attribute profile for the input image a. Note that all melanocytes present a significant elongation, however some post processing is required in order to eliminate porous structures on the background. Most noisy regions are not area-stable, then the area-stable elongation $\Phi(X_t)$ appears suitable for the segmentation of this kind of objects, as shown in (d). This example demonstrates the use of our area-stable elongation in order to enhance elongated objects with respect to a noisy background. Using this feature image, the melanocyte segmentation becomes an easy task.

connected operator since quasi-flat zones do not create new contours in the image, and it is also auto-dual since bright, dark and intermediate gray level regions are processed at the same time. The natural application of this method is the segmentation of objects with a given attribute or shape. However, we have shown that its application domain is wider. For example, when this controlled propagation is computed for each pixel on a



(a) Input image b

(b) Ground truth

(c) Maximal elongation image: $\max\{E(X_t)\}$

(d) Maximal area-sable elongation image: $\max\{\Phi(X_t)\}$

Figure 7.17: Feature images using the global maximum in the attribute profile for the input image b. Note that all melanocytes present a significant elongation, however some post processing is required in order to eliminate porous structures on the background. Most noisy regions are not area-stable, then the area-stable elongation $\Phi(X_t)$ appears suitable for the segmentation of this kind of objects, as shown in (d). This example demonstrates the use of our area-stable elongation in order to enhance elongated objects with respect to a noisy background. Using this feature image, the melanocyte segmentation becomes an easy task.

pilot image, input-adaptive SE can be defined and shape features can be assessed. The main advantage of our approach is that no size parameter is required in order to determine the appropriate region.

The main drawback is the chaining effect due to transition regions, *i.e.* paths with gradual transitions connect different regions of the image in the same λ-flat zone. As consequence, the propagation can reach

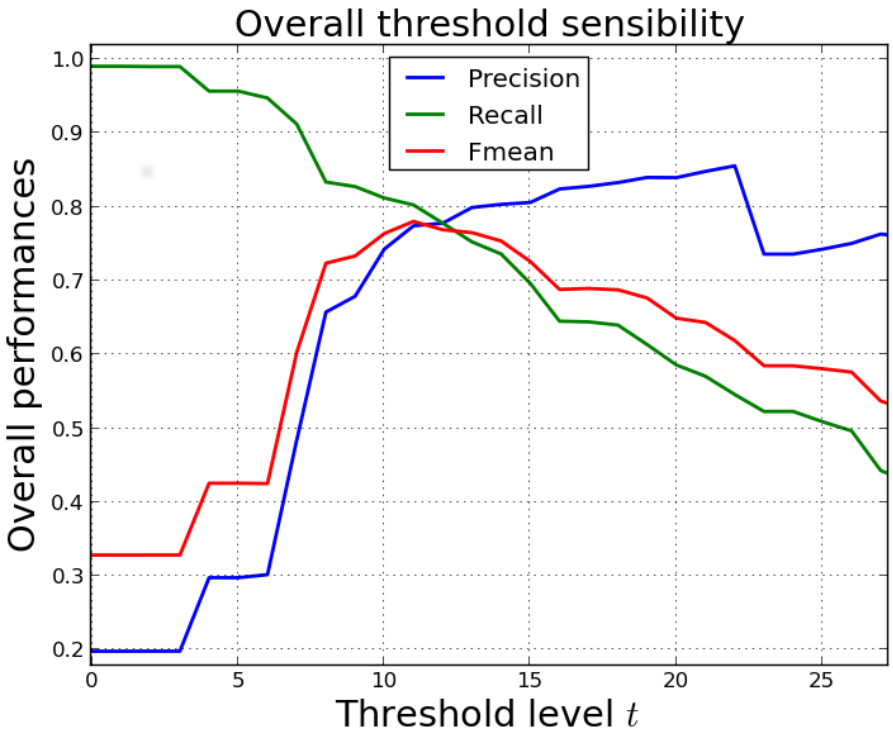
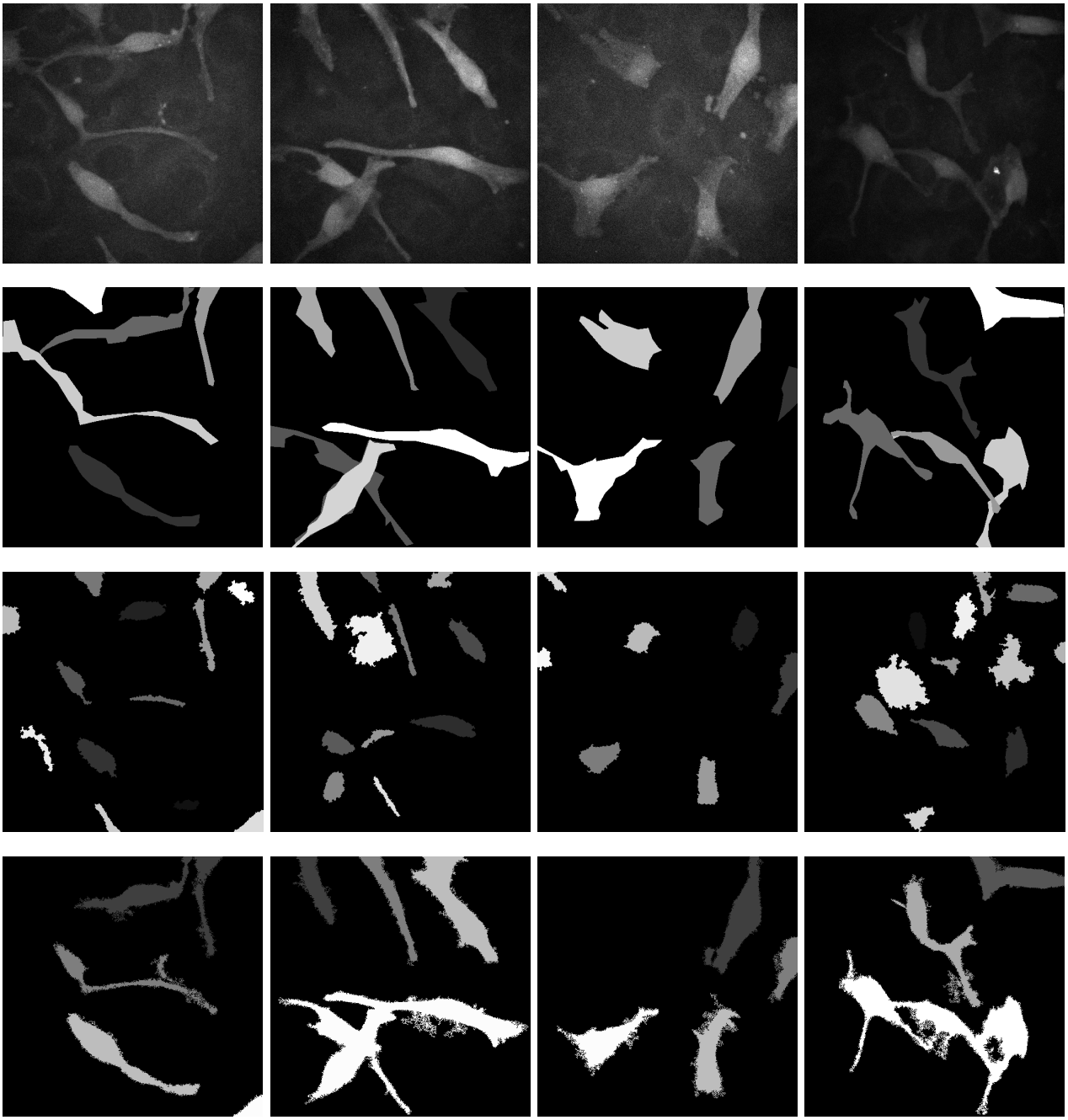


Figure 7.18: Overall sensibility curves: threshold to eliminate objects with low area-stable elongation. This parameter is not critical since several values produce similar results. It is noteworthy that thresholds between 7 and 16 produce an overall f_{mean} over 70%. In our experiments we have used a threshold equal to 11 for all images.

different objects through these paths, even for low λ values. In such a case, two solutions are recommended: to use a filtered version of the input image as pilot image and to use another type of propagation.

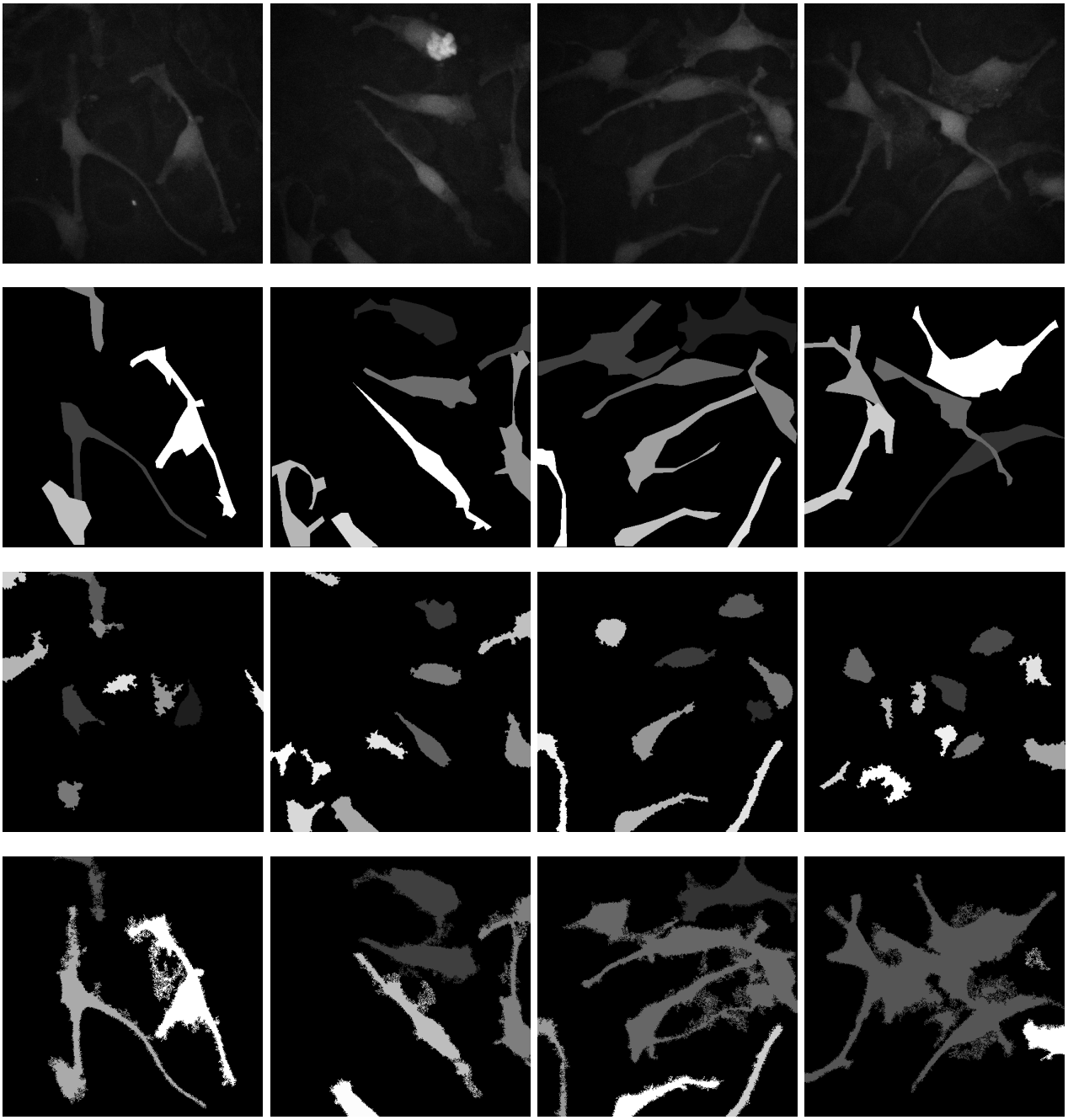
Therefore, we have proposed an extension of this method: the analysis of attribute profiles on the image. For this analysis another connected hierarchical partition is used. Images are firstly represented as component trees using threshold decomposition. Then, the attribute profile is analyzed and important events are recorded. In particular, two well-known attributes are used: geodesic elongation and area. These two attributes have been combined to define a new attribute: the *area-stable elongation*. The behavior of this new attribute in relation to noise, blur and geometrical distortions is discussed. The global maximum of this attribute is computed for each pixel of the input image and a feature image is built. Such image is a spatial partition where objects of interest can be easily extracted. This method can be interpreted as an extension of MSER favoring objects of a given shape. A difference with the classical MSER is that only the global maximum of the attribute profile is chosen, thus only the most stable and elongated region is kept. This new attribute has been successfully used in a cosmetic application aiming at segmenting melanocytes cells that appear as bright and elongated structures in multiphoton images of engineered skin. Standard methods may fail because melanocytes are low contrasted and noisy. It has been proven that better segmentations are obtained providing a prior knowledge about cells shape. One of the method limitations, common to all methods based on threshold decomposition, is that it can only segment CC present in the component tree.

As general remark, the present chapter confirms the interest of attribute-based operators for image filtering and segmentation. In future works, other interesting attributes such as porosity and tortuousity will be studied. Additionally, extensions to higher dimensional data (color, multi-spectral or 3D) will be analyzed. In such a case, other metrics should be used to define quasi-flat propagation rules and ordering in the component tree.



(a) Image a (b) Image b (c) Image c (d) Image d

Figure 7.19: Segmentation of melanocytes using area-stable elongation. First row: input image; second row: ground truth; third row: MSER; fourth row: our segmentation result. Note that our method presents much better results than MSER for all images.



(e) Image e (f) Image f (g) Image g (h) Image h

Figure 7.20: Segmentation of melanocytes using area-stable elongation (continuation). First row: input image; second row: ground truth; third row: MSER; fourth row: our segmentation result. Note that our method presents much better results than MSER for all images.

8 Conclusions and Perspectives

8.1 Résumé

Dans ce chapitre, nous présenterons les conclusions de cette thèse sur l'analyse sémantique de nuages de points 3D dans le milieu urbain. Nous exposerons les principales contributions de chaque chapitre ainsi que les perspectives à venir pour la recherche dans ce domaine.

8.2 Conclusions

Most important cities in the world have very detailed urban plans of streets and public spaces. These plans contain information about sidewalks, roads and urban furniture such as street lights, traffic signs, trees, bus shelters, among others. Recently, multi-source information such as topographical surveying, crowd-sourcing, analysis from satellite images, urban photos, 3D point clouds, among others, is being integrated into Geographical Information Systems (GIS), simplifying the management of urban information. Nowadays, several local authorities, national mapping agencies and private companies are including 3D information in urban maps. 3D data open a wide range of applications such as urban planning, cultural heritage documentation, virtual tourism, itinerary planning, marketing, navigation systems, video games, itinerary planning for soft mobility, accessibility diagnoses, among others.

Recent technological advances make laser scanning an accurate and productive solution for the acquisition of 3D urban data required for those maps. Compared to the first 3D scanning systems 30 years ago, current laser scanners are cheaper, faster and provide more accurate and denser 3D point clouds. For including such data into a 3D map, the usual pipeline includes transforming 3D points into surfaces or geometric primitives for subsequent analysis. These transformations are usually carried out by manually assisted approaches, leading to time consuming procedures, unsuitable for large scale applications. Manual object extraction from urban scenes is difficult and tedious, and existing semi-automatic methods may not be sufficiently precise nor robust, then exhaustive manual corrections are necessary. In that sense, automatic methods for semantic analysis of 3D urban data are required.

This Ph.D. thesis, entitled: "Semantic analysis of 3D point clouds from urban environments: ground, facades, urban objects and accessibility", has been developed at MINES ParisTech in the Center for Mathematical Morphology (CMM) in Fontainebleau (France) under the supervision of Dr. Beatriz Marcotegui Iturmendi. We aim at developing automatic methods to process 3D point clouds from urban laser scanning. Our methods are based on elevation images, mathematical morphology and supervised learning. The development of accurate and fast algorithms in this domain is the main contribution of this thesis. We have focused on a complete 3D urban analysis method including six main steps: i) Filtering and preprocessing; ii) ground segmentation and accessibility analysis; iii) facade segmentation, iv) object detection; v) object segmentation; vi) object classification.

This thesis has been developed in the framework of TerraMobilita project: *"3D mapping of roads and urban public space, accessibility and soft mobility"* (presented in Chapter 1). Then, our experiments have been conducted in agreement with project requirements and applications on urban mobility, accessibility analysis, and public space management. In that sense, we have also worked on the integration of our results into a large-scale production chain. Our results are exported as 3D point clouds for visualization and modeling purposes and as shapefiles for integration in any GIS.

First, let us summarize and discuss the contributions of this thesis. Then, perspectives for future work will be presented.

8.3 Contributions of this thesis

Before studying methods for 3D semantic urban analysis, we have reviewed (in Chapter 2) the different laser scanning technologies used in urban environments. Additionally, we have presented public 3D databases in the

state of the art and we have found that few 3D urban databases are publicly available and manual annotations are rarely found in the literature. Therefore, as part of this thesis and in the framework of TerraMobilita project, we have collaborated in the creation, annotation and publication of several 3D urban databases[1] (Serna et al., 2014b; Brédif et al., 2014) as well as in the definition of evaluation protocols using 2D and 3D manual annotations (Serna and Marcotegui, 2013b; Brédif et al., 2014). Additionally, we have co-organized, in cooperation with the National French Mapping Agency (IGN), an international contest[2] aiming at benchmarking semantic analysis methods working on 3D dense urban data (Vallet et al., 2014).

3D datasets are delivered as long lists of (x, y, z) coordinates, possibly with attributes such as intensity, color, GPS time, among others. Points are usually listed in scan line order, which is not suitable for efficient processing. A suitable data structure is not only required to inspect and to visualize 3D information, but also to process it conveniently. Several data structures such as elevation images, triangulation, meshing, octrees and k-D trees have been proposed in the state of the art. The choice of the best data structure is application dependent and it is possible to combine some of them to get better results in specific tasks such as visualization, filtering, segmentation and classification. In Chapter 3, an overview on these 3D data structures has been presented. We have proposed the use of elevation images since they are convenient structures to visualize and to process data using all the large collection of image processing tools, in particular mathematical morphology. Projecting 3D information to images implies a reduction in the amount of data to be processed with respect to the input 3D point cloud. Besides, neighborhood relationships in the elevation image are easily computed. In general, processing an elevation image using image processing techniques is much faster than processing the 3D point cloud directly. Although the idea of deriving elevation images from 3D point clouds is not new, this thesis confirms their usefulness in the development of accurate and fast urban analysis methods.

From the processing point of view, ground segmentation is one of the most important steps in urban semantic analysis since all the urban entities are located on it. In Chapter 4, we have proposed a straightforward but robust method for accessibility analysis in urban environments. Ground is segmented using the quasi-flat (denoted by the symbol λ-flat) zones labeling algorithm, which allows to segment the ground even in the presence of access ramps, speed humps and other non–flat structures (Hernández and Marcotegui, 2009a). Next, gradient information is used in order to detect elevation discontinuities on the ground. Then, curb candidates are selected, close curbs are reconnected using Bézier curves and semantic information. Finally, geometric characterization is carried out and accessibility is defined based on international standards. This constitutes one of the most attractive contributions of this thesis due to its social impact since urban accessibility affects not only disabled persons but also old people, children and pregnant women. In the framework of the *United Nations convention on the rights of persons with disabilities*, local authorities are required to guarantee accessibility in public spaces in order to reduce social exclusion, low employment and limited education of people concerned by accessibility. Thus, it is very important to be able to make large-scale accessibility diagnoses in urban environments. One of our publications on this topic (Serna and Marcotegui, 2013b) has been awarded with the U. V. Helava Award[3] for the 2013 best paper in the *International Society for Photogrammetry and Remote Sensing* (ISPRS Journal volumes 75-86). The Jury justified this award as follows:

> "This paper addresses the problem of detecting navigable routes for wheelchairs in urban areas based on curb detection from mobile laser scanner point clouds. A key scientific contribution noted by the Jury is a new method for providing continuity of extracted curb lines using Bézier curves. The Jury was impressed with the results and felt that the social impact of their very practically-focused research could be wide-reaching in society as the future demand for accessibility information will likely be very high."

Once the ground is segmented, all remaining structures are considered as facades and objects. Discrimination between them is important because facades delimit the end of public space while urban objects define the obstacle map required for itinerary planning. In Chapter 5, we have proposed several automatic methods to segment facades. In our experiments, facades are high, vertical and elongated structures on the elevation image. Our facade segmentation methods are based on geometrical and geodesic constraints. Given the urban and architectural constraints of our databases, most of parameters have been set intuitively. Three facade segmentation approaches have been proposed: reconstruction by dilation from markers, attribute controlled reconstruction from markers, and segmentation based on the maximal elongation image (without markers). The method based on reconstruction by dilation is the fastest one since it is based on iterative geodesic dilations in order to get the entire facade. Its main problem is that objects connected to the facade are reconstructed

[1] For further information, the reader is encouraged to visit: http://cmm.ensmp.fr/~serna/downloads.html
[2] For further information, the reader is encouraged to visit: http://data.ign.fr/benchmarks/UrbanAnalysis/
[3] For further information, the reader is encouraged to visit: http://www.isprs.org/society/awards/helava/2013.aspx

in the facade mask. For overcoming such problem, we have proposed a second method based on attribute controlled reconstruction using geodesic elongation. Since connected objects usually appear at low height and reduce the global facade elongation, this method offers better results than the first one. In general, methods based on facade markers are strongly influenced by the markers extraction method. The main drawback is that bad located markers produce errors since they may reconstruct non-facade objects. For this reason, we have proposed a more robust method avoiding the use of facade markers. In this method, only the elongation and its evolution over the height decomposition of the scene are analyzed. This method is based on the maximal elongation image computed from 3D decomposition. This third method has proved to produce the best results, but its implementation is slower. However, it remains suitable for large-scale applications since processing takes only a few tens of seconds for an acquisition of several hundreds of meters, using a non-optimized implementation. The selection of the best facade segmentation method remains application dependent. It should be a trade off between quality results and computational cost. In the case of a large-scale application, where time constraints are less strict, the most accurate method should be preferred. Independently of the method, facade segmentation result is used to segment city blocks. City blocks are considered as the biggest semantic entities in the urban environment. Their segmentation is carried out using influence zones. Each city block can be processed separately and each individual result joined at the end of the analysis, reducing memory requirements and allowing parallelization.

In Chapter 6, we have presented a semantic analysis of 3D urban objects based on mathematical morphology and supervised learning. The focus is automatic detection, segmentation and classification of urban objects from 3D laser scanning data. Our automatic method generates object hypotheses as discontinuities and bumps on the ground. Then, connected objects are segmented in order to assign a unique identifier to each individual object. Our method is proven to be robust to noise since small and isolated structures are eliminated using morphological filters. Our main under- and over-segmentation problems are due to parked motorcycles and pedestrians walking too close to cars, which may not be correctly separated. After segmentation, objects are classified in several categories using an SVM approach with geometrical and contextual features. Our geometrical features can be adapted to any XYZ point cloud. Thus, the classification can be easily generalized, *i.e.* training on a database and testing on another one (as presented in Section 6.8.4). This is a significant advantage because the model learned from a database can be applied to another one, even acquired by a different acquisition system, without the tedious manual annotation. One of the main drawbacks processing 3D urban data using elevation images is that high objects may occlude lower objects located below them. That is why we propose an alternative segmentation strategy using two slices. In the lower slice, objects are processed as aforementioned, while in the upper slice, a rule–based method has been proposed in order to segment trees, poles and off-ground objects. It is obvious that processing two slices is more expensive than processing only one elevation image. Therefore, two slices are only used in databases containing several trees or other high objects occluding objects below them. In particular, this strategy has been successfully applied in TerraMobilita/iQmulus database. Other databases such as Paris-rues-Vaugirard-Madame and Paris-rue-Soufflot do not contain trees, thus processing by slices has not been required. In the case of Ohio database, most trees are present in the east side of the city, which has been acquired by aerial laser scanning. In that case, lower tree parts are not visible and processing by slices is not justified.

Finally, we have presented in Chapter 7 several methodological contributions to mathematical morphology. We have developed powerful attribute-based operators useful in a wide range of applications such as: attribute controlled reconstruction, adaptive mathematical morphology, feature extraction, filtering and segmentation. The natural application of these methods in the urban semantic analysis is the segmentation of elongated objects such as curbs and facades, presented in Chapters 4 and 5, respectively. Additionally, we have presented other applications such as the segmentation of elongated cells in an industrial context. As general remark, this last contribution confirms the interest of attribute-based operators for image filtering and segmentation.

8.4 Perspectives

Our methods have been qualitative and quantitative tested in several databases from the state of the art (Ohio, Enschede and Paris-rue-Soufflot) and from TerraMobilita project (Paris-rue-Madame and TerraMobilita/iQmulus). Even if our methods have presented good results and have outperformed state of the art methods, it is noteworthy that several improvements should be done before developing a mature application. Our main problem, common to all methods in the literature, is due to large occluded regions. Several scans of the same zone, as those produced by velodyne sensors, could reduce this problem.

Several under- and over-segmentation problems have been also pointed out. Up to now, we have only used the

spatial information available in the point cloud. A possible solution can include shape/texture analysis to help deciding whether an object should be re-segmented. Moreover, additional features such as laser intensity and texture could improve performances in detection, segmentation and classification steps. In that sense, new tools need to be developed in order to deal with 3D textured data (Angulo, 2011; Angulo and Velasco-Forero, 2014). For example, we can define a filter which output not only depends on the texture but also on the neighborhood depth. This tool, using the same philosophy of bilateral filters (Paris et al., 2009), could be used to re-segment urban objects without merging information from background or other objects since they are at different depths and have different textures.

Our contributions to mathematical morphology have confirmed the interest of attribute-based operators in filtering and segmentation tasks. In future works, other interesting attributes such as porosity and tortuousity will be studied. Additionally, extensions to higher dimensional data (color, multi-spectral or 3D) will be analyzed. In such a case, other metrics should be used to define quasi-flat propagation rules and ordering in the component tree.

Our approaches have been implemented in a research prototype, mainly based on Morph-M library (CMM, 2013), the image processing library of the CMM. The library allows easy prototyping but it is not intended to be fast. In spite of this non-optimized implementation, our current methods are suitable for large-scale applications and are currently used by TerraMobilita project partners, as they are much faster than any manual-assisted method. For example, manual annotation speed was approximately 50 meters per hour in our Paris-rue-Madame (Serna et al., 2014b) and TerraMobilita/iQmulus (Brédif et al., 2014) databases. In a city like Paris, with 1700 km of streets, approximately 4 years will be required for a complete manual annotation. As it has been proved in this thesis, our processing takes only a few tens of seconds for an acquisition of several hundreds of meters, providing accurate results. Currently, the optimization of our base operators (erosion, dilation, opening, reconstruction, watershed, among others) is under development at CMM in order to bring optimized operators for real-time and big-data problems. Software (hierarchical queues,structuring elements decomposition, among others) and hardware (SIMD-Single Instruction Multiple Data and parallelization) optimizations are being integrated in SMIL library (Faessel and Bilodeau, 2013) and will be integrated in our future developments.

Bibliography

1Spatial, 2014. Elyx 3D. http://1spatial.com/products-services/elyx/elyx-gis-platform/elyx-3d (Last accessed: October 10, 2014).

ADA, 2010. 2010 ADA Standards for Accessible Design. U.S. Department of Justice, http://www.ada.gov/2010ADAstandards_index.htm (Last accessed: May 20, 2012).

Aijazi, A. K., Checchin, P., Trassoudaine, L., 2013. Segmentation Based Classification of 3D Urban Point Clouds: A Super-Voxel Based Approach with Evaluation. Remote Sensing 5 (4), 1624–1650.

Alexander, C., Tansey, K., Kaduk, J., Holland, D., Tate, N. J., 2010. Backscatter coefficient as an attribute for the classification of full-waveform airborne laser scanning data in urban areas. ISPRS Journal of Photogrammetry and Remote Sensing 65 (5), 423–432.

Anguelov, D., Taskarf, B., Chatalbashev, V., Koller, D., Gupta, D., Heitz, G., Ng, A., 2005. Discriminative learning of Markov random fields for segmentation of 3D scan data. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR. Vol. 2. pp. 169–176.

Angulo, J., 2011. Morphological Bilateral Filtering and Spatially-Variant Adaptive Structuring Functions. In: Proceedings of the 10th International Symposium on Mathematical Morphology (ISMM 2011). pp. 212–223.

Angulo, J., Velasco-Forero, S., 2014. Riemannian mathematical morphology. Pattern Recognition Letters 47, 93–101, advances in Mathematical Morphology.

Archivideo, 2014. Archivideo, from France in 3D to the 3D world. http://www.archivideo.com/ (Last accessed: October 10, 2014).

Arp, H., Griesbach, J., Burns, 1982. Mapping in tropical forests: a new approach using the Laser APR. Photogrammetric Engineering and Remote Sensing 48 (1), 91–100.

Avci, M., Akyurek, Z., 2004. A Hierarchical Classificaton of Landsat Tm Imagery for Landcover Mapping. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXV-B4. pp. 511–516.

Ayres, T., Kelkar, R., 2006. Sidewalk potential trip points: A method for characterizing walkways. International Journal of Industrial Ergonomics 36 (12), 1031–1035.

Bab-Hadiashar, A., Gheissari, N., 2006. Range image segmentation using surface selection criterion. IEEE Transactions on Image Processing 15 (7), 2006–2018.

Beger, R., Gedrange, C., Hecht, R., Neubert, M., 2011. Data fusion of extremely high resolution aerial imagery and LiDAR data for automated railroad centre line reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing 66 (6), 40–51, advances in LIDAR Data Processing and Applications.

Bentley, J. L., 1975. Multidimensional binary search trees used for associative searching. Communications of the ACM 18 (9), 509–517.

Beucher, S., 1987. Traffic Spatial Measurements Using Video Image Processing. Intelligent Robots and Computer Vision 848, 648–655.

Beucher, S., 2007. Numerical residues. Image and Vision Computing 25 (4), 405–415.

Beucher, S., Meyer, F., 1993. The morphological approach to segmentation: the watershed transformation. In: Dougherty, E. R. (Ed.), Mathematical Morphology in Image Processing. Marcel Dekker, New York, Ch. 12, pp. 433–481.

Boulaassal, H., Grussenmeyer, P., Tarsha-kurdi, F., 2007. Automatic segmentation of building facades using terrestrial laser data. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVI-3/W52. pp. 65–70.

Boyer, K., Mirza, M., Ganguly, G., 1994. The Robust Sequential Estimator: a general approach and its application to surface organization in range data. IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (10), 987–1001.

Brédif, M., Vallet, B., Serna, A., Marcotegui, B., Paparoditis, N., 2014. TerraMobilita/iQmulus urban point cloud classification benchmark. In: IQmulus workshop on Processing Large Geospatial Data. iQmulus/TerraMobilita contest, pp. 1–6, http://data.ign.fr/benchmarks/UrbanAnalysis/ (Last accessed: June 26, 2014).

Breen, E. J., Jones, R., 1996. Attribute Openings, Thinnings, and Granulometries. Computer Vision and Image Understanding 64 (3), 377–389.

Bulatov, D., Rottensteiner, F., Schulz, K., 2012. Context-based urban terrain reconstruction from images and videos. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. I-3. pp. 185–190.

CapDigital, 2009. Terra Numerica : La numérisation du patrimoine urbain. http://www.competitivite.gouv.fr/spip.php?article636 (Last accessed: Septembre 11, 2014).

CapDigital, 2014. TerraMobilita:3D mapping of roads and urban public space, accessibility and soft mobility. http://cmm.ensmp.fr/TerraMobilita/ (Last accessed: Septembre 11, 2014).

CASQY, 2014. Saint-Quentin-en-Yvelines communauté d'agglomération. http://www.saint-quentin-en-yvelines.fr/ (Last accessed: October 10, 2014).

Chaperon, T., Goulette, F., 2001. Extracting cylinders in full 3D data using a random sampling method and the Gaussian image. In: Proceedings of the Vision Modeling and Visualization Conference, VMV-01. Aka GmbH, Stuttgart, Germany.

CMM, 2013. Morph-M : Image processing software specialized in Mathematical Morphology. MINES ParisTech, CMM - Center for Mathematical Morphology, http://morphm.ensmp.fr (Last accessed: October 8, 2014).

CoE LaSR, 2013. Centre of Excellence in Laser Scanning Research. Finnish Geodetic Institut, http://www.fgi.fi/coelasr/ (Last accessed: December 16, 2013).

Cramer, M., 2010. The DGPF test on digital aerial camera evaluation - overview and test design. Photogrammetrie - Fernerkundung - Geoinformation 2, 73–82.

Cretu, A. M., Petriu, E., Patry, G., 2006. Neural–network–based models of 3D objects for virtualized reality: a comparative study. IEEE Transactions on Instrumentation and Measurement 55 (1), 99–111.

Cyclomedia, 2014. Cyclomedia: smart imagery solutions. http://www.cyclomedia.com/en/ (Last accessed: October 10, 2014).

Delaunay, B., 1934. Sur la sphère vide. Otdelenie Matematicheskikh i Estestvennykh Nauk 7, 793–800.

Demantke, J., Mallet, C., David, N., Vallet, B., 2010. Dimensionality based scale selection in 3D LiDAR point clouds. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVIII-5/W12. pp. 97–102.

Denis, E., Burck, R., Baillard, C., 2010. Towards road modelling from terrestrial laser points. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII-3A, 293–298.

Deschaud, J.-E., 2010. Traitements de nuages de points denses et modélisation 3D d'environnements par système mobile LiDAR/Caméra. Ph.D. thesis, MINES ParisTech.

Douglas, D., Peucker, T., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. The Canadian Cartographer 10 (2), 112–122.

Douillard, B., Underwood, J., Kuntz, N., Vlaskine, V., Quadros, A., Morton, P., Frenkel, A., 2011. On the segmentation of 3D LIDAR point clouds. In: IEEE International Conference on Robotics and Automation, ICRA'11. pp. 2798–2805.

Duda, R. O., Hart, P. E., Stork, D. G., 2000. Pattern Classification, 2nd Edition. Wiley Interscience.

Earthmine, 2014. Earthmine: complete solutions for 3D streel level imagery. http://www.earthmine.com/html/home.html (Last accessed: October 10, 2014).

Edelsbrunner, H., Mücke, E. P., 1994. Three-dimensional alpha shapes. ACM Transactions on Graphics 13, 43–72.

Faessel, M., Bilodeau, M., Mar. 2013. SMIL: Simple Morphological Image Library. In: Séminaire Performance et Généricité, LRDE. Villejuif, France, http://cmm.ensmp.fr/~faessel/smil (Last accessed: October 8, 2014).

Ferguson, D., Darms, M., Urmson, C., Kolski, S., 2008. Detection, prediction, and avoidance of dynamic obstacles in urban environments. In: IEEE Intelligent Vehicles Symposium. pp. 1149–1154.

Figueroa, P., Londoño, E., Prieto, F., Boulanger, P., Borda, J., Restrepo, D., 2006. Experiencias virtuales con piezas del Museo del Oro de Colombia. Tech. rep., RENATA: Red Nacional Acadï£¡mica de Tecnología Avanzada, http://www.renata.edu.co/index.php/ciencias-sociales/ (Last accessed: December 20, 2010).

Forssen, P.-E., 2007. Maximally Stable Colour Regions for Recognition and Matching. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007. pp. 1–8.

Forssen, P.-E., Lowe, D., 2007. Shape Descriptors for Maximally Stable Extremal Regions. In: IEEE 11th International Conference on Computer Vision, ICCV 2007. pp. 1–8.

Franchi, G., Angulo, J., 2014. Spatially-variant area openings for reference-driven adaptive contour preserving filtering. Tech. rep., MINES ParisTech.

Fritzke, B., 1995. A growing Neural Gas Network learns topologies. In: Advances in Neural Information Processing Systems. Vol. 7. MIT Press, pp. 625–632.

Gang, L., Guangshun, S., 2010. Procedural Modeling of Urban Road Network. In: International Forum on Information Technology and Applications (IFITA 2010). Vol. 1. pp. 75–79.

García, J., Amaral, P., Marrón, M., Mazo, M., Bastos Filho, T., 2010. Proposal for an Ambient Assisted Wheelchair (A2W). In: IEEE International Symposium on Industrial Electronics (ISIE 2010). pp. 2325–2330.

Geoautomation, 2014. GeoAutomation Âů Next Generation Surveying. http://www.geoautomation.com/ (Last accessed: October 10, 2014).

Gerke, M., Xiao, J., 2013. Supervised and unsupervised MRF based 3D scene classification in multiple view airborne oblique images. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. II-3/W3. pp. 25–30.

Gerke, M., Xiao, J., 2014. Fusion of airborne laserscanning point clouds and images for supervised and unsupervised scene classification. ISPRS Journal of Photogrammetry and Remote Sensing 87 (0), 78–92.

Golovinskiy, A., Kim, V. G., Funkhouser, T., 2009. Shape-based recognition of 3D point clouds in urban environments. In: 12th IEEE International Conference on Computer Vision. pp. 2154–2161.

Gonzalez, R. C., Woods, R. E., 2006. Digital Image Processing, 3rd Edition. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Google, I., 2014a. Google earth. https://www.google.fr/intl/fr/earth/ (Last accessed: October 10, 2014).

Google, I., 2014b. Google Trekker.
URL http://www.google.fr/maps/about/behind-the-scenes/streetview/treks

Gordon, R., Rangayyan, R. M., 1984. Feature enhancement of film mammograms using fixed and adaptive neighborhoods. Applied Optics 23 (4), 560–564.

Gorte, B., 2007. Planar feature extraction in terrestrial laser scans using gradient based range image segmentation. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVI-3/W52. pp. 173–177.

Goulette, F., Nashashibi, F., Ammoun, S., Laurgeau, C., 2006a. An Integrated on–Board Laser Range Sensing System for On–the–Way City and Road Modelling. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (A), 3–5.

Goulette, F., Nashashibi, F., Ammoun, S., Laurgeau, C., 2006b. An integrated on-board laser range sensing system for On-the-Way City and Road Modelling. The ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVI-1, 1–6.

Grazzini, J., Soille, P., 2008. Adaptive Morphological Filtering Using Similarities Based on Geodesic Time. In: Coeurjolly, D., Sivignon, I., Tougne, L., Dupont, F. (Eds.), Discrete Geometry for Computer Imagery. Vol. 4992. Springer, pp. 519–528.

Grigillo, D., Kanjir, U., 2012. Urban object extraction from digital surface model and digital aerial images. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. I-3. pp. 215–220.

Hambrusch, S., He, X., Miller, R., 1994. Parallel Algorithms for Gray-Scale Digitized Picture Component Labeling on a Mesh–Connected Computer. Journal of Parallel and Distributed Computing 20 (1), 56–68.

Hammoudi, K., 2011. Contributions to the 3D city modeling. Ph.D. thesis, Université Paris-Est.

Hernández, J., 2009. Analyse morphologique d'images pour la modélisation d'environnements urbains. Ph.D. thesis, MINES ParisTech.

Hernández, J., Marcotegui, B., 2009a. Filtering of artifacts and pavement segmentation from mobile LiDAR data. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVIII-3/W8. pp. 329–333.

Hernández, J., Marcotegui, B., 2009b. Morphological segmentation of building façade images. In: IEEE International Conference on Image Processing, ICIP'09. pp. 4029–4032.

Hernández, J., Marcotegui, B., 2009c. Point cloud segmentation towards urban ground modeling. In: The 5th GRSS/ISPRS Joint Urban Remote Sensing Event (URBAN 2009). Shangai, China, pp. 1–5.

Hervieu, A., Soheilian, B., 2013a. Road Side Detection and Reconstruction Using LIDAR Sensor. In: IEEE Intelligent Vehicles Symposium. pp. 23–26.

Hervieu, A., Soheilian, B., 2013b. Semi-Automatic Road/Pavement Modeling using Mobile Laser Scanning. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. II-3/W3. pp. 31–36.

Hoover, A., 1994. University of South Florida (USF) Range Image Database. http://marathon.csee.usf.edu/range/DataBase.html (Last accessed: May 15, 2014).

Hoover, A., Jean-baptiste, G., Jiang, X., Flynn, P. J., Bunke, H., Goldgof, D. B., Bowyer, K., Eggert, D. W., Fitzgibbon, A., Fisher, R. B., 1996. An explerimental comparison of range image segmentation algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (7), 673–689.

Hubel, D. H., Wiesel, T. N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of Physiology 160 (1), 106–154.

IGN, 2013. IGN - Geospatial and Terrestrial Imagery. IGN French National Mapping Agency, http://isprs.ign.fr/home_en.htm (Last accessed: October 28, 2013).

IGN, 2014a. Géoportail: Le portail des territoires et des citoyens. http://www.geoportail.gouv.fr/accueil (Last accessed: October 10, 2014).

IGN, 2014b. IGN: Institut national de l'information géographique et forestière. http://www.ign.fr/ (Last accessed, October 10, 2014).

INSPIRE, 2007. INSPIRE: Infrastructure for Spatial Information in Europe. Environment Directorate-General of the European Commission. http://inspire.ign.fr/directive/presentation (Last accessed: October 10, 2014).

ISO, 2008. ISO 7176-5: Wheelchairs–Part 5: Determination of dimensions, mass and manoeuvring space. ISO–International Organization for Standardization.

ISPRS, 2013. The ISPRS data set collection. ISPRS International Society for Photogrammetry and Remote Sensing, http://www.isprs.org/data/ (Last accessed: October 28, 2013).

Jähne, B., 2005. Digital Image Processing. Springer.

Jones, R., 1999. Connected Filtering and Segmentation Using Component Trees. Computer Vision and Image Understanding 75 (3), 215–228.

Kammel, S., Ziegler, J., Pitzer, B., Werling, M., Gindele, T., Jagzent, D., Schröder, J., Thuy, M., Goebl, M., Hundelshausen, F. v., Pink, O., Frese, C., Stiller, C., 2008. Team AnnieWAY's autonomous system for the 2007 DARPA Urban Challenge. Journal of Field Robot. 25 (9), 615–639.

Kimmel, R., Zhang, C., Bronstein, A., Bronstein, M., 2011. Are MSER Features Really Interesting? IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (11), 2316–2320.

Kohonen, T., 2001. Self-Organizing Maps, 3rd Edition. Information Sciences. Springer-Verlag, Heidelberg.

Kohonen, T., Nieminen, I., Honkela, T., 2009. On the quantization error in SOM vs. VQ: a critical and systematic study. In: Príncipe, J., Miikkulainen, R. (Eds.), Advances in Self–Organizing Maps. Vol. 5629 of Lecture Notes in Computer Science. Springer Berlin/Heidelberg, German, pp. 133–144.

Krabill, W. B., Collins, J. G., Link, L. E., Swift, R. N., Butler, M. L., 1984. Airborne laser topographic mapping results. Photogrammetric Engineering and Remote Sensing 50, 685–694.

Lafarge, F., Mallet, C., 2012. Creating large-scale city models from 3D-point clouds: a robust approach with hybrid representation. International Journal of Computer Vision 99 (1), 69–85.

Lantuéjoul, C., Beucher, S., 1981. On the use of the geodesic metric in image analysis. Journal of Microscopy 121 (1), 39–49.

Lantuéjoul, C., Maisonneuve, F., 1984. Geodesic methods in quantitative image analysis. Pattern Recognition 17 (2), 177–187.

Lari, Z., Habib, A., 2014. An adaptive approach for the segmentation and extraction of planar and linear/cylindrical features from laser scanning data. ISPRS Journal of Photogrammetry and Remote Sensing 93 (1), 192–212.

Lerallut, R., Decencière, E., Meyer, F., 2007. Image filtering using morphological amoebas. Image and Vision Computing 25 (4), 395–404.

Lindengerber, J., 1989. Test results of laser profiling for topographic terrain survey. In: Proceedings 42nd Photogrammetric week. Wichmann Verlag, pp. 25–39.

Litman, R., Bronstein, A., Bronstein, M., 2012. Stable volumetric features in deformable shapes. Computers & Graphics 36 (5), 569–576, Shape Modeling International Conference, SMI 2012.

LoiHandicap, 2005. Loi 2005-102: "Pour l'égalité des droits et des chances, la participation et la citoyenneté des personnes handicapées". http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000809647&dateTexte=&categorieLien=id (Last accessed: Septembre 11, 2014).

Mallet, C., Bretar, F., Roux, M., Soergel, U., Heipke, C., 2011. Relevance assessment of full-waveform LiDAR data for urban area classification. ISPRS Journal of Photogrammetry and Remote Sensing 66 (6), 71–84.

Mallet, C., Bretar, F., Soergel, U., 2008. Analysis of Full-Waveform LiDAR data for classification of urban areas. Photogrammetrie - Fernerkundung - Geoinformation (PFG) 5, 337–349.

Maragos, P., Vachier, C., 2009. Overview of adaptive morphology: Trends and perspectives. In: 16th IEEE International Conference on Image Processing (ICIP2009). pp. 2241–2244.

Maragos, P., Ziff, R., 1990. Threshold superposition in morphological image analysis systems. IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (5), 498–504.

Marshall, D., Lukacs, G., Martin, R., 2001. Robust Segmentation of Primitives from Range Data in the Presence of Geometric Degeneracy. IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (3), 304–314.

Martinetz, T., Berkovich, S., Schulten, K., 1993. 'Neural-gas' network for vector quantization and its application to time-series prediction. IEEE Transactions on Neural Networks 4 (4), 558–569.

Matas, J., Chum, O., Urban, M., Pajdla, T., 2004. Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing 22 (10), 761–767, British Machine Vision Computing, BMVC 2002.

Matheron, G., 1975. Random Sets and Integral Geometry. John Wiley & Sons, New York.

Meagher, D. J. R., 1980. Octree encoding: a new technique for the representation, manipulation and display of arbitrary 3-D objects by computer. Tech. rep., Electrical and Systems Engineering Department Rensseiaer Polytechnic Institute Image Processing Laboratory.

Menkens, C., Sussmann, J., Al-Ali, M., Breitsameter, E., Frtunik, J., Nendel, T., Schneiderbauer, T., 2011. EasyWheel - A Mobile Social Navigation and Support System for Wheelchair Users. In: Eighth International Conference on Information Technology: New Generations ITNG 2011. pp. 859–866.

Meyer, F., 1998. From connected operators to levelings. In: Mathematical Morphology and its Applications to Image and Signal Processing. Vol. 12 of Computational Imaging and Vision. Kluwer Academic Publishers, pp. 191–198.

Meyer, F., Vachier, C., 2002. Image segmentation based on viscous flooding simulation. In: Talbot, H., Beare, R. (Eds.), Proceedings of the 5th International Symposium on Mathematical Morphology (ISMM 2002). Sydney, Australy, pp. 69–77.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L., 2005. A Comparison of Affine Region Detectors. International Journal of Computer Vision 65 (1-2), 43–72.

Morard, V., Decencière, E., Dokládal, P., 2011a. Geodesic attributes thinnings and thickenings. In: Proceedings of the 10th International Symposium on Mathematical Morphology (ISMM 2011). Springer-Verlag, pp. 200–211.

Morard, V., Decencière, E., Dokládal, P., 2011b. Region Growing Structuring Elements and New Operators based on their Shape. In: International conference on Signal and Image Processing (SIP'11). ACTA Press, pp. 1–8.

Morard, V., Decencière, E., Dokládal, P., 2013. Efficient geodesic attribute thinnings based on the barycentric diameter. Journal of Mathematical Imaging and Vision 46 (1), 128–142.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 66 (3), 247–259.

Munoz, D., Vandapel, N. D., Hebert, M., 2009. Onboard contextual classification of 3-D point clouds with learned high-order Markov Random Fields. In: IEEE International Conference on Robotics and Automation, ICRA '09. pp. 2009–2016.

Na, S., Xumin, L., Yong, G., 2010. Research on k–means clustering algorithm: an improved k-means clustering algorithm. In: Third International Symposium on Intelligent Information Technology and Security Informatics (IITSI). pp. 63–67.

Nagao, M., Matsuyama, T., Ikeda, Y., 1979. Region extraction and shape analysis in aerial photographs. Computer Graphics and Image Processing 10 (3), 195–223.

Niemeyer, J., Rottensteiner, F., Soergel, U., 2014. Contextual classification of lidar data and building object detection in urban areas. ISPRS Journal of Photogrammetry and Remote Sensing 87 (0), 152–165.

Nistér, D., Stewénius, H., 2008. Linear Time Maximally Stable Extremal Regions. In: European Conference on Computer Vision, ECCV 2008. Vol. 5303 of LNCS. Springer Berlin Heidelberg, pp. 183–196.

Nüchter, A., Lingemann, K., 2011. Robotic 3D Scan Repository. Jacobs University Bremen gGmbH and University of Osnabrück, http://kos.informatik.uni-osnabrueck.de/3Dscans/ (Last accessed: December 16, 2013).

Okabe, A., Boots, B., Sugihara, K., 1992. Spatial tessellations. Concepts and applications of Voronoi diagrams. J. Wiley and Sons.

Otsu, N., 1979. A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man and Cybernetics 9 (1), 62–66.

Ouzounis, G., Pesaresi, M., Soille, P., 2012. Differential Area Profiles: Decomposition Properties and Efficient Computation. IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (8), 1533–1548.

Owechko, Y., Medasani, S., Korah, T., 2010. Automatic recognition of diverse 3-D objects and analysis of large urban scenes using ground and aerial LiDAR sensors. In: Conference on Lasers and Electro-Optics (CLEO) and Quantum Electronics and Laser Science Conference (QELS). pp. 16–21.

PagesJaunes, 2007. Pages Jaunes: Ville en 3D. http://v3d.pagesjaunes.fr/ (Last accessed: October 10, 2014).

Paparoditis, N., Papelard, J.-P., Cannelle, B., Devaux, A., Soheilian, B., David, N., Houzay, E., 2012. Stereopolis II: A multi–purpose and multi-sensor 3D mobile mapping system for street visualisation and 3D metrology. Revue Française de Photogrammétrie et de Télédétection 200 (1), 69–79.

Paris, 2014. Mairie de Paris: Direction de la voirie et des déplacements (DVD). http://www.paris.fr/politiques/organigramme-des-directions-services/direction-de-la-voirie-et-des-deplacements-dvd/p5385 (Last accessed: October 10, 2014).

Paris, S., Kornprobst, P., Tumblin, J., Durand, F., 2009. Foundations and Trends in Computer Graphics and Vision. Vol. 4. Ch. Bilateral Filtering: Theory and Applications, pp. 1–73.

Parker, J. R., 2010. Algorithms for image processing and computer vision. John Wiley & Sons.

Perona, P., Malik, J., 1990. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (7), 629–639.

Pesaresi, M., Benediktsson, J., 2001. A new approach for the morphological segmentation of high-resolution satellite imagery. IEEE Transactions on Geoscience and Remote Sensing 39 (2), 309–320.

Pinoli, J.-c., Debayle, J., 2009. General Adaptive neighborhood mathematical morphology. In: 16th IEEE International Conference on Image Processing (ICIP'09). pp. 2249–2252.

Poggio, T., Shelton, C. R., 1999. Machine learning, machine vision, and the brain. AI Magazine 20 (3), 37–56.

Poreba, M., Goulette, F., 2012a. Assessing the Accuracy of Land-Based Mobile Laser Scanning Data. Geomatics and Environmental Engineering 6 (3), 73–81.

Poreba, M., Goulette, F., 2012b. RANSAC algorithm and elements of graph theory for automatic plane detection in 3D point clouds. Archives of Photogrammetry, Cartography and Remote Sensing 24, 301–310.

Pu, S., Rutzinger, M., Vosselman, G., Elberink, S. O., 2011. Recognizing basic structures from mobile laser scanning data for road inventory studies. ISPRS Journal of Photogrammetry and Remote Sensing 66 (6), 28–39.

Rashid, O., Dunabr, A., Fisher, S., Rutherford, J., 2010. Users Helping Users: User Generated Content to Assist Wheelchair Users in an Urban Environment. In: Ninth International Conference on Mobile Business and 2010 Ninth Global Mobility Roundtable (ICMB-GMR-2010). pp. 213–219.

Roerdink, J. B. T. M., 2009. Adaptivity and group invariance in mathematical morphology. In: Proceedings of the International Conference on Image Processing (ICIP'09). Cairo, Egypt, pp. 2253–2256.

Rutzinger, M., Pratihast, A. K., Oude Elberink, S. J., Vosselman, G., 2011. Tree modelling from mobile laser scanning data-sets. The Photogrammetric Record 26 (135), 361–372.

Salembier, P., Serra, J., 1995. Flat zones filtering, connected operators and filters by reconstruction. IEEE Transactions on Image Processing 4 (8), 1153–1160.

Salembier, P., Wilkinson, M. H. F., 2009. Connected operators. IEEE Signal Processing Magazine 26 (6), 136–157.

Schmitt, M., Preteux, F., 1986. A new mathematical morphological algorithm: r,h maxima and r,h minima. Application to X ray tomographs, N.M.R., angiography. In: Second image processing symposium: image processing, computer generated images, technology and applications.

Schnabel, R., Wessel, R., Wahl, R., Klein, R., 2008. Shape recognition in 3D point clouds. In: The 16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision. Union Agency-Science Press, pp. 1–8.

Serna, A., 2011. Modelado 3D de tumores cerebrales empleando endoneurosonografía y redes neuronales artificiales. Master's thesis, National university of Colombia - Manizales.

Serna, A., Hernández, J., Marcotegui, B., 2012. Adaptive parameter tuning for morphological segmentation of building facade images. In: Proceedings of the 20th European Signal Processing Conference (EUSIPCO2012).

Serna, A., Marcotegui, B., 2012. Classification 3D d'objets urbains à partir des données terrestres à balayage laser. In: 35ème journée ISS France. École des Mines de Paris, Paris, France, pp. 1–2.

Serna, A., Marcotegui, B., 2013a. Attribute controlled reconstruction and adaptive mathematical morphology. In: 11th International Symposium on Mathematical Morphology (ISMM 2013). Uppsala, Sweden, pp. 205–216.

Serna, A., Marcotegui, B., 2013b. Urban accessibility diagnosis from mobile laser scanning data. ISPRS Journal of Photogrammetry and Remote Sensing 84, 23–32.

Serna, A., Marcotegui, B., jul 2014. Detection, segmentation and classification of 3D urban objects using mathematical morphology and supervised learning. ISPRS Journal of Photogrammetry and Remote Sensing 93, 243–255.

Serna, A., Marcotegui, B., Decencière, E., Baldeweck, T., Pena, A.-M., Brizion, S., 2014a. Segmentation of elongated objects using attribute profiles and area stability: Application to melanocyte segmentation in engineered skin. Pattern Recognition Letters 47, 172–182.

Serna, A., Marcotegui, B., Goulette, F., Deschaud, J.-E., et al., 2014b. Paris-rue-Madame database: a 3D mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods. In: 3rd International Conference on Pattern Recognition, Applications and Methods ICPRAM 2014. pp. 1–4.

Serra, J., 1982. Image analysis and mathematical morphology. Vol. 1. Academic Press, Orlando, FL, USA.

Serra, J., 1988. Image analysis and mathematical morphology: theoretical advances. Vol. 2. Academic Press, London.

Serra, J., 1993. The "Centre de Morphologie Mathématique": An overview. In: 2nd international symposium on mathematical morphology (ISMM 93).

Serra, J., 1998. Connectivity on Complete Lattices. Journal of Mathematical Imaging and Vision 9 (3), 231–251.

Serra, J., 2005. Viscous Lattices. Journal of Mathematical Imaging and Vision 22, 269–282.

Serra, J., Salembier, P., 1993. Connected operators and pyramids. SPIE Image Algebra and Morphological Image Processing 2030, 65–76.

Serra, J., Soille, P., 1994. Mathematical Morphology and Its Applications to Image Processing. Kluwer Academic Publishers.

Sevcik, C., Studnicka, N., 2006. Documentation of complex facades and city modelling through the combination of Laserscanning and photogrammetry. Tech. rep., GeoDATA IT mbh and RIEGL Laser Measurement Systems.

Shao, H., Svodoba, T., Gool, L. V., 2003. ZuBuD - Zurich Building Database for Image Based recognition. Tech. Rep. 260, Computer Vision Lab, ETH Zurich.

Shapovalov, R., Velizhev, A., Barinova, O., 2010. Non-associative Markov networks for 3D point cloud classification. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII (3A), 103–108.

Shih, F. Y., Cheng, S., 2004. Adaptive mathematical morphology for edge linking. Information Sciences 167 (1), 9–21.

Siegemund, J., Pfeiffer, D., Franke, U., Förstner, W., 2010. Curb Reconstruction using Conditional Random Fields. In: IEEE Intelligent Vehicles Symposium (IV). IEEE Computer Society, pp. 203–210.

Soille, P., 2003. Morphological image analysis: principles and applications. Springer-Verlag, Secaucus, NJ, USA.

Soille, P., 2008. Constrained connectivity for hierarchical image decomposition and simplification. IEEE transactions on pattern analysis and machine intelligence 30 (7), 1132–1145.

Soille, P. J., Ansoult, M. M., 1990. Automated basin delineation from digital elevation models using mathematical morphology. Signal Processing 20 (2), 171–182.

Sozialhelden, 2012. wheelmap. SOZIALHELDEN e.V., Berlin, Germany, http://wheelmap.org/en/ (Last accessed: September 9, 2014).

Talbot, H., Appleton, B., 2007. Efficient complete and incomplete path openings and closings. Image and Vision Computing 25 (4), 416–425.

Teboul, O., Simon, L., Koutsourakis, P., Paragios, N., 2010. Segmentation of building facades using procedural shape priors. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR2010. pp. 3105–3112.

Teeravech, K., Nagai, M., Honda, K., Dailey, M., 2014. Discovering repetitive patterns in facade images using a RANSAC style algorithm. ISPRS Journal of Photogrammetry and Remote Sensing 92 (1), 38–53.

Trimble, 2014a. Trimble 3D Laser Scanning. Trimble Navigation Limited, http://www.trimble.com/3d-laser-scanning/ (Last accessed at: June 18, 2014).

Trimble, 2014b. Trimble RealWorks: 3D scanning software. Trimble Navigation Limited, http://www.trimble.com/3d-laser-scanning/software.aspx (Last accessed at: June 13, 2014).

Trimble, 2014c. Trimble TX8 laser scanner. Trimble Navigation Limited, http://www.trimble.com/3d-laser-scanning/tx8.aspx?dtID=overview& (Last accessed at: June 18, 2014).

UN, 2007. United Nations Convention on the Rights of Persons with disabilities. http://www.un.org/disabilities/convention/conventionfull.shtml (Last accessed: September 9, 2014).

Valero, S., Chanussot, J., Benediktsson, J. A., Talbot, H., Waske, B., 2010. Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images. Pattern Recognition Letters 31 (10), 1120–1127.

Vallet, B., Brédif, M., Serna, A., Marcotegui, B., Paparoditis, N., 2014. TerraMobilita/iQmulus Urban Point Cloud Analysis Benchmark. Computers & Graphics. (Submitted on September 4, 2014).

Velizhev, A., Shapovalov, R., Schindler, K., 2012. Implicit shape model for object detection in 3D point clouds. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. I-3. pp. 179–184.

Velodyne, 2012. The HDL-32E Velodyne LiDAR sensor. Velodyne Lidar 2012, http://velodynelidar.com/lidar/hdlproducts/hdl32e.aspx (Last accessed: December 16, 2013).

Verly, J., Delanoy, R., 1993. Adaptive mathematical morphology for range imagery. IEEE Transactions on Image Processing 2 (2), 272–275.

Vincent, L., 1993. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. IEEE Transactions on Image Processing 2 (2), 176–201.

Vincent, L., 1994. Morphological area openings and closings for grey-scale images. In: Proceedings of the Workshop: shape in picture. Springer, Driebergen, The Netherlands, pp. 197–208.

Voronoi, G., 1908. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Journal für die Reine und Angewandte Mathematik 133, 97–178.

Vosselman, G., Maas, H.-G., 2010. Airborne and Terrestrial Laser Scanning. Whittles Publishing.

Vosselman, G., Zhou, L., 2009. Detection of curbstones in airborne laser scanning data. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVIII-3/W8. pp. 11–116.

Weinmann, M., Jutzi, B., Mallet, C., 2013. Feature relevance assessment for the semantic interpretation of 3D point cloud data. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. II-5/W5. pp. 313–318.

Weinmann, M., Jutzi, B., Mallet, C., 2014. Semantic 3D scene interpretation: A framework combining optimal neighborhood size selection with relevant features. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. II(3). pp. 181–188.

Wendt, P., Coyle, E., Gallagher, N. J., 1986. Stack filters. IEEE Transactions on Acoustics, Speech and Signal Processing 34 (4), 898–911.

Werghi, N., Fisher, R., Robertson, C., Ashbrook, A., 1998. Modelling Objects Having Quadric Surfaces Incorporating Geometric Constraints. In: Proceedings of 5th European Conference on Computer Vision ECCV'98. Springer-Verlag, pp. 185–201.

Wikipedia, 2014. City blocks. http://en.wikipedia.org/w/index.php?title=City_block&oldid=617153038 (Last accessed: October 8, 2014).

Zhou, L., Vosselman, G., 2012. Mapping curbstones in airborne and mobile laser scanning data. International Journal of Applied Earth Observation and Geoinformation 18, 293–304.

Zhu, X., Zhao, H., Liu, Y., Zhao, Y., Zha, H., 2010. Segmentation and classification of range image from an intelligent vehicle in urban environment. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010). pp. 1457–1462.

Analyse sémantique de nuages de points 3D dans le milieu urbain : sol, façades, objets urbains et accessibilité

Résumé : Les plus grandes villes au monde disposent de plans 2D très détaillés des rues et des espaces publics. Ces plans contiennent des informations relatives aux routes, trottoirs, façades et objets urbains tels que, entre autres, les lampadaires, les panneaux de signalisation, les poteaux, et les arbres. De nos jours, certaines autorités locales, agences nationales de cartographie et sociétés privées commencent à adjoindre à leurs cartes de villes des informations en 3D, des choix de navigation et d'accessibilité. En comparaison des premiers systèmes de scanning en 3D d'il y a 30 ans, les scanners laser actuels sont moins chers, plus rapides et fournissent des nuages de points 3D plus précis et plus denses. L'analyse de ces données est difficile et laborieuse, et les méthodes semi-automatiques actuelles risquent de ne pas être suffisamment précises ni robustes. C'est en ce sens que des méthodes automatiques pour l'analyse urbaine sémantique en 3D sont nécessaires.

Cette thèse constitue une contribution au domaine de l'analyse sémantique de nuages de points en 3D dans le cadre d'un environnement urbain. Nos méthodes sont basées sur les images d'élévation et elles illustrent l'efficacité de la morphologie mathématique pour développer une chaîne complète de traitement en 3D, incluant 6 étapes principales : i) filtrage et pré-traitement ; ii) segmentation du sol et analyse d'accessibilité ; iii) segmentation des façades ; iv) détection d'objets ; v) segmentation d'objets ; vi) classification d'objets. De plus, nous avons travaillé sur l'intégration de nos résultats dans une chaîne de production à grande échelle. Ainsi, ceux-ci ont été incorporés en tant que "shapefiles" aux Systèmes d'Information Géographique et exportés en tant que nuages de points 3D pour la visualisation et la modélisation.

Nos méthodes ont été testées d'un point de vue qualitatif et quantitatif sur plusieurs bases de données issues de l'état de l'art et du projet TerraMobilita. Nos résultats ont montré que nos méthodes s'avèrent précises, rapides et surpassent les travaux décrits par la littérature sur ces mêmes bases. Dans la conclusion, nous abordons également les perspectives de développement futur.

Mots clés : Morphologie Mathématique, Traitement d'Image, Analyse Urbaine en 3D, Accessibilité Urbaine, Analyse Sémantique, Segmentation, Classification.

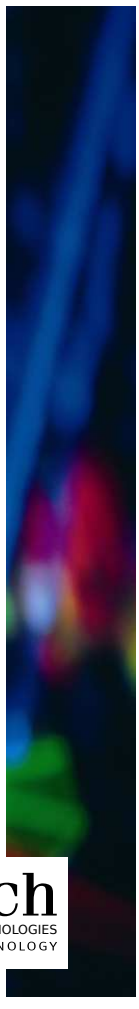
Semantic analysis of 3D point clouds from urban environments: ground, facades, urban objects and accessibility.

Abstract: Most important cities in the world have very detailed 2D urban plans of streets and public spaces. These plans contain information about roads, sidewalks, facades and urban objects such as lampposts, traffic signs, bollards, trees, among others. Nowadays, several local authorities, national mapping agencies and private companies have began to consider justifiable including 3D information, navigation options and accessibility issues into urban maps. Compared to the first 3D scanning systems 30 years ago, current laser scanners are cheaper, faster and provide more accurate and denser 3D point clouds. Urban analysis from these data is difficult and tedious, and existing semi-automatic methods may not be sufficiently precise nor robust. In that sense, automatic methods for 3D urban semantic analysis are required.

This thesis contributes to the field of semantic analysis of 3D point clouds from urban environments. Our methods are based on elevation images and illustrate how mathematical morphology can be exploited to develop a complete 3D processing chain including six main steps: i) filtering and preprocessing; ii) ground segmentation and accessibility analysis; iii) facade segmentation, iv) object detection; v) object segmentation; and, vi) object classification. Additionally, we have worked on the integration of our results into a large-scale production chain. In that sense, our results have been exported as 3D point clouds for visualization and modeling purposes and integrated as shapefiles into Geographical Information Systems (GIS).

Our methods have been qualitative and quantitative tested in several databases from the state of the art and from TerraMobilita project. Our results show that our methods are accurate, fast and outperform other works reported in the literature on the same databases. Conclusions and perspectives for future work are discussed as well.

Keywords: Mathematical Morphology, Image Processing, 3D Urban Analysis, Urban Accessibility, Semantic Analysis, Segmentation, Classification.



MINES ParisTech

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY